

INFORMATION PROCESSING DEVICE AND METHOD, RECORDING MEDIUM AND PROGRAM

Publication number: JP2003242176 (A)

Publication date: 2003-08-29

Inventor(s): SAITO MARI, YAMAMOTO NORIYUKI +

Applicant(s): SONY CORP +

Classification:

- international: G06F17/30; G06F17/30; (IPC1-7): G06F17/30

- European: G06F17/30T1E; G06F17/30T5S

Application number: JP20020147225 20020522

Priority number(s): JP20020147225 20020522; JP20010379611 20011213

Also published as:

US2003140369 (A1)

US7289982 (B2)

Abstract of JP 2003242176 (A)

PROBLEM TO BE SOLVED: To present related information according to the degree of interest of a user. ; SOLUTION: In step S101 and 102, the number of times of transmission (mail reply) (exchange frequency) is counted for every mail opposite party and the mail opposite parties are sorted in the order of higher frequency of transmission. In step S103 and S104, the mail opposite parties in upper ranks of the exchange frequency are extracted, and the weight of words contained in the mails with the extracted opposite parties is corrected. For instance, the mail exchange with Mr. A to be the top rank of the exchange frequency is a topic A and the weight of words b, c and f contained in the topic A is multiplied by a weighting factor k set in advance to be corrected. As a result, the weight of the words contained in the topic that the user feels more interests is corrected. This invention can be applied to a personal computer. ; COPYRIGHT: (C)2003, JPO

図31



Data supplied from the **espacenet** database — Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2003-242176
(P2003-242176A)

(43) 公開日 平成15年8月29日 (2003.8.29)

(51) Int.Cl. ⁷ G 0 6 F 17/30	識別番号 3 8 0 1 7 0 2 1 0 3 4 0 3 5 0	F 1 C 0 6 F 17/30	7-エコード (参考) 3 8 0 D 5 B 0 7 0 1 7 0 A 2 1 0 A 3 4 0 A 3 5 0 C
審査請求 未請求 請求項の数69 O L (全 34 頁)			
(21) 出願番号	特願2002-147225(P2002-147225)	(71) 出願人	000002185 ソニー株式会社 東京都品川区北品川6丁目7番35号
(22) 出願日	平成14年5月22日(2002.5.22)	(72) 発明者	斎藤 真里 東京都品川区北品川6丁目7番35号 ソニー株式会社内
(31) 優先権主張番号	特願2001-379511(P2001-379511)	(73) 発明者	山本 則行 東京都品川区北品川6丁目7番35号 ソニー株式会社内
(32) 優先日	平成13年12月13日(2001.12.13)	(74) 代理人	100082131 弁理士 榎本 義雄
(33) 優先権主張国	日本 (J P)		
最終頁に続			

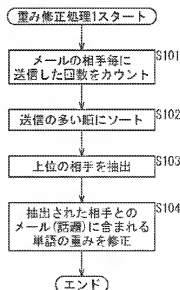
(54) 【発明の名称】 情報処理装置および方法、記録媒体、並びにプログラム

(57) 【要約】

【目的】 ユーザの興味の高合いに応じて関連情報を提示するようにする。

【構成】 ステップS101、102において、メールの相手毎に、送信（返信）した回数（やりとりの頻度）がカウントされ、送信の頻度の多い順にソートされる。ステップS103、S104において、やりとりの頻度が上位のメールの相手が抽出され、抽出された相手とのメールに含まれる単語の重みが修正される。例えば、やりとりの頻度が1位のAさんとのメールのやりとりが話題Aであり、その話題に含まれる単語b、c、dの重みに、予め設定されている重み係数kが乗算されて修正される。これにより、ユーザがより興味を感じている話題に含まれる単語の重みが修正される。本発明は、パーソナルコンピュータに適用できる。

図31



【特許請求の範囲】

【請求項1】 既存の文書情報および前記の文書情報に関連する関連情報を提示する情報処理装置において、前記既存の文書情報から第1の特徴語を抽出するとともに、前記所定の文書情報から第2の特徴語を抽出する抽出手段と、

前記抽出手段により抽出された、前記第1の特徴語の第1の重みを算出するとともに、前記第2の特徴語の第2の重みを算出する重み算出手段と、

所定の条件に基づいて、前記重み算出手段により算出された前記第1の重みを修正する重み修正手段と、前記抽出手段により抽出された前記第1の特徴語、および前記重み修正手段により修正された前記第1の重みに基づいて、前記既存の文書情報に付随する前記関連情報を取得する取得手段と、

前記第2の特徴語に基づいて、前記所定の文書情報に関連する前記既存の文書情報を検索する検索手段と、前記検索手段により検索された前記既存の文書情報に対応付けられている前記関連情報の表示を制御する表示制御手段とを備えることを特徴とする情報処理装置。

【請求項2】 前記取得手段は、前記既存の文書情報における、複数の前記第1の特徴語の中から、前記第1の重みに基づいて重要語を選択し、前記重要語に関連する前記関連情報を取得することを特徴とする請求項1に記載の情報処理装置。

【請求項3】 前記検索手段は、前記第1の重みと前記第2の重みから類似度を算出し、前記類似度に基づいて、前記所定の文書情報に対応する前記既存の文書情報を検索することを特徴とする請求項1に記載の情報処理装置。

【請求項4】 前記既存の文書情報および前記所定の文書情報から属性情報を抽出する属性情報抽出手段をさらに備えることを特徴とする請求項1に記載の情報処理装置。

【請求項5】 前記属性情報抽出手段により抽出された前記属性情報に基づいて、1または複数の前記既存の文書情報をグループ化して文書情報群を作成するグループ化手段をさらに備え、

前記抽出手段は、前記文書情報群から前記第1の特徴語を抽出し、

前記重み算出手段は、前記第1の特徴語の前記第1の重みを算出し、

前記取得手段は、前記第1の特徴語および前記第1の重みに基づいて、前記文書情報群に対応付ける前記関連情報を取得し、

前記検索手段は、前記第2の特徴語に基づいて、前記所定の文書情報に対応する前記文書情報群を検索することを特徴とする請求項1に記載の情報処理装置。

【請求項6】 前記重み修正手段は、前記属性情報抽出手段により抽出された前記属性情報に基づいて、前記第1

の重みを修正することを特徴とする請求項4に記載の情報処理装置。

【請求項7】 前記重み修正手段は、特定の宛先への送信または前記特定の宛先からの受信の頻度に応じて、前記第1の重みを修正することを特徴とする請求項1に記載の情報処理装置。

【請求項8】 前記重み修正手段は、前記情報処理装置の使用者が作成した前記既存の文書情報から抽出された前記第1の特徴語の前記第1の重みを修正することを特徴とする請求項1に記載の情報処理装置。

【請求項9】 前記重み修正手段は、他の前記既存の文書情報から抽出された、前記第1の特徴語と同じ特徴語の重みを、前記第1の重みと同様にさらに修正することを特徴とする請求項1に記載の情報処理装置。

【請求項10】 前記重み修正手段は、前記第1の特徴語の種類に基づいて、前記第1の重みを修正することを特徴とする請求項1に記載の情報処理装置。

【請求項11】 前記既存の文書情報の前記第1の特徴語に関連付けて、前記取得手段により取得された前記関連情報を用いてデータベースを構築するデータベース構築手段をさらに備え、

前記検索手段は、前記第2の特徴語に基づいて、前記データベース構築手段により構築された前記データベースから、前記所定の文書情報に関連する前記既存の文書情報を検索することを特徴とする請求項1に記載の情報処理装置。

【請求項12】 所定のイベントの発生を検知する検知手段をさらに備え、

前記所定の文書情報は、前記検知手段により検知された前記所定のイベントの発生に対応する文書であることを特徴とする請求項1に記載の情報処理装置。

【請求項13】 前記既存の文書情報および前記所定の文書情報は電子メールであることを特徴とする請求項1に記載の情報処理装置。

【請求項14】 所定のデータ構造で構成された単語群からなる、単語間の関係を示すデータベースである分野辞書を作成する分野辞書作成手段をさらに備え、

前記重み修正手段は、前記分野辞書作成手段により作成された前記分野辞書に基づいて、前記重み算出手段により算出された前記第1の重みを修正することを特徴とする請求項1に記載の情報処理装置。

【請求項15】 前記分野辞書は、前記単語群が、各前記単語の属する分野によってグループ化され、記録されているデータベースであることを特徴とする請求項1に記載の情報処理装置。

【請求項16】 前記分野辞書は、前記単語群が、前記単語間の関係に基づいて階層構造に配置され、記録されているデータベースであることを特徴とする請求項1に記載の情報処理装置。

【請求項17】 前記分野辞書は、前記単語群が、各前

記単語に割り当てられた重要度に基づいてリスト化されたデータベースであることを特徴とする請求項1に記載の情報処理装置。

【請求項18】 前記分野辞書は、既存の辞書データベースに含まれる前記単語群および前記単語群が形成するデータ構造を含むことを特徴とする請求項1に記載の情報処理装置。

【請求項19】 前記分野辞書は、所定のWebページデータに含まれる前記単語群および前記単語群が形成するデータ構造を含むことを特徴とする請求項1に記載の情報処理装置。

【請求項20】 前記分野辞書は、以前に前記検索手段により用いられた前記第2の特徴語のリストに含まれる前記単語群および前記単語群が形成するデータ構造を含むことを特徴とする請求項1に記載の情報処理装置。

【請求項21】 前記重み修正手段は、前記分野辞書に含まれる前記単語に対応する前記第1の特徴語の前記第1の重みを修正することを特徴とする請求項1に記載の情報処理装置。

【請求項22】 既存の文書情報および所定の文書情報に関連する関連情報を提示する情報処理装置の情報処理方法において、前記既存の文書情報から第1の特徴語を抽出するとともに、前記所定の文書情報から第2の特徴語を抽出する抽出ステップと、

前記抽出ステップの処理により抽出された、前記第1の特徴語の第1の重みを算出するとともに、前記第2の特徴語の第2の重みを算出する重み算出ステップと、所定の条件に基づいて、前記重み算出ステップの処理により算出された前記第1の重みを修正する重み修正ステップと、

前記抽出ステップの処理により抽出された前記第1の特徴語、および前記重み修正ステップの処理により修正された前記第1の重みに基づいて、前記既存の文書情報に対応付ける前記関連情報の取得を制御する取得制御ステップと、

前記第2の特徴語に基づいて、前記所定の文書情報に関連する前記既存の文書情報を検索する検索ステップと、前記検索ステップの処理により検索された前記既存の文書情報に対応付けられている前記関連情報の表示を制御する表示制御ステップとを含むことを特徴とする情報処理方法。

【請求項23】 前記取得制御ステップは、前記既存の文書情報における、複数の前記第1の特徴語の中から、前記第1の重みに基づいて重要語を選択し、前記重要語に関連する前記関連情報の取得を制御することを特徴とする請求項22に記載の情報処理方法。

【請求項24】 前記検索ステップは、前記第1の重みと前記第2の重みから類似度を算出し、前記類似度に基づいて、前記所定の文書情報に対応する前記既存の文書

情報を検索することを特徴とする請求項22に記載の情報処理方法。

【請求項25】 前記既存の文書情報および前記所定の文書情報から属性情報を抽出する属性情報抽出ステップをさらに含み、

前記重み修正ステップは、前記属性情報抽出ステップの処理により抽出された前記属性情報に基づいて、前記第1の重みを修正することを特徴とする請求項22に記載の情報処理方法。

【請求項26】 前記属性情報抽出ステップの処理により抽出された前記属性情報に基づいて、1または複数の前記既存の文書情報をグループ化して文書情報群を作成するグループ化ステップをさらに含むことを特徴とする請求項22に記載の情報処理方法。

【請求項27】 前記重み修正ステップは、特定の宛先への送信または前記特定の宛先からの受信の頻度に応じて、前記第1の重みを修正することを特徴とする請求項22に記載の情報処理方法。

【請求項28】 前記重み修正ステップは、前記第1の特徴語の種類に基づいて、前記第1の重みを修正することを特徴とする請求項22に記載の情報処理方法。

【請求項29】 前記既存の文書情報の前記第1の特徴語に関連付けて、前記取得制御ステップの処理により取得が制御された前記関連情報を用いてデータベースを構築するデータベース構築ステップをさらに含むことを特徴とする請求項22に記載の情報処理方法。

【請求項30】 所定のデータ構造で構成された単語群からなる、単語間の関係を示すデータベースである分野辞書を作成する分野辞書作成ステップをさらに含み、前記重み修正ステップは、前記分野辞書作成ステップの処理により作成された前記分野辞書に基づいて、前記重み算出ステップの処理により算出された前記第1の重みを修正することを特徴とする請求項22に記載の情報処理方法。

【請求項31】 前記分野辞書は、前記単語群が、各前記単語の属する分野によってグループ化され、記録されているデータベースであることを特徴とする請求項22に記載の情報処理方法。

【請求項32】 前記分野辞書は、前記単語群が、前記単語間の関係に基づいて階層構造に配置され、記録されているデータベースであることを特徴とする請求項22に記載の情報処理方法。

【請求項33】 前記分野辞書は、前記単語群が、各前記単語に割り当てられた重要度に基づいてリスト化されたデータベースであることを特徴とする請求項22に記載の情報処理方法。

【請求項34】 前記分野辞書は、既存の辞書データベースに含まれる前記単語群および前記単語群が形成するデータ構造を含むことを特徴とする請求項22に記載の情報処理方法。

【請求項35】 前記分野辞書は、所定のWEBページデータに含まれる前記単語群および前記単語群が形成するデータ構造を含むことを特徴とする請求項22に記載の情報処理方法。

【請求項36】 前記分野辞書は、以前に前記検索ステップの処理により用いられた前記第2の特徴語のリストに含まれる前記単語群および前記単語群が形成するデータ構造を含むことを特徴とする請求項22に記載の情報処理方法。

【請求項37】 前記重み修正ステップは、前記分野辞書に含まれる前記単語に対応する前記第1の特徴語の前記第1の重みを修正することを特徴とする請求項22に記載の情報処理方法。

【請求項38】 既存の文書情報および所定の文書情報に関連する関連情報を提示する情報処理装置を制御するプログラムであって、前記既存の文書情報から第1の特徴語を抽出するとともに、前記所定の文書情報から第2の特徴語を抽出する抽出ステップと、

前記抽出ステップの処理により抽出された、前記第1の特徴語の第1の重みを算出するとともに、前記第2の特徴語の第2の重みを算出する重み算出ステップと、

所定の条件に基づいて、前記重み算出ステップの処理により算出された前記第1の重みを修正する重み修正ステップと、

前記抽出ステップの処理により抽出された前記第1の特徴語、および前記重み修正ステップの処理により修正された前記第1の重みに基づいて、前記既存の文書情報に対応付ける前記関連情報の取得を制御する取得制御ステップと、

前記第2の特徴語に基づいて、前記所定の文書情報に関連する前記既存の文書情報を検索する検索ステップと、前記検索ステップの処理により検索された前記既存の文書情報に対応付けられている前記関連情報の表示を制御する表示制御ステップとを含むことを特徴とするコンピュータが読み取り可能なプログラムが記録されている記録媒体。

【請求項39】 前記取得制御ステップは、前記既存の文書情報における、複数の前記第1の特徴語の中から、前記第1の重みに基づいて重要語を選択し、前記重要語に関連する前記関連情報の取得を制御することを特徴とする請求項38に記載の記録媒体。

【請求項40】 前記検索ステップは、前記第1の重みと前記第2の重みから類似度を算出し、前記類似度に基づいて、前記所定の文書情報に対応する前記既存の文書情報を検索することを特徴とする請求項38に記載の記録媒体。

【請求項41】 前記既存の文書情報および前記所定の文書情報から属性情報を抽出する属性情報抽出ステップをさらに含み、

前記重み修正ステップは、前記属性情報抽出ステップの処理により抽出された前記属性情報に基づいて、前記第1の重みを修正することを特徴とする請求項38に記載の記録媒体。

【請求項42】 前記属性情報抽出ステップの処理により抽出された前記属性情報に基づいて、1または複数の前記既存の文書情報をグループ化して文書情報群を作成するグループ化ステップをさらに含むことを特徴とする請求項38に記載の記録媒体。

【請求項43】 前記重み修正ステップは、特定の宛先への送信または前記特定の宛先からの受信の頻度に応じて、前記第1の重みを修正することを特徴とする請求項38に記載の記録媒体。

【請求項44】 前記重み修正ステップは、前記第1の特徴語の種類に基づいて、前記第1の重みを修正することを特徴とする請求項38に記載の記録媒体。

【請求項45】 前記既存の文書情報の前記第1の特徴語に関連付けて、前記取得制御ステップの処理により取得が制御された前記関連情報を用いてデータベースを構築するデータベース構築ステップをさらに含むことを特徴とする請求項38に記載の記録媒体。

【請求項46】 所定のデータ構造で構成された単語群からなる、単語間の関係を示すデータベースである分野辞書を作成する分野辞書作成ステップをさらに含み、前記重み修正ステップは、前記分野辞書作成ステップの処理により作成された前記分野辞書に基づいて、前記重み算出ステップの処理により算出された前記第1の重みを修正することを特徴とする請求項38に記載の記録媒体。

【請求項47】 前記分野辞書は、前記単語群が、各前記単語の属する分野によってグループ化され、記録されているデータベースであることを特徴とする請求項38に記載の記録媒体。

【請求項48】 前記分野辞書は、前記単語群が、前記単語間の関係に基づいて階層構造に配置され、記録されているデータベースであることを特徴とする請求項38に記載の記録媒体。

【請求項49】 前記分野辞書は、前記単語群が、各前記単語に割り当てられた重要性に基づいてリスト化されたデータベースであることを特徴とする請求項38に記載の記録媒体。

【請求項50】 前記分野辞書は、既存の辞書データベースに含まれる前記単語群および前記単語群が形成するデータ構造を含むことを特徴とする請求項38に記載の記録媒体。

【請求項51】 前記分野辞書は、所定のWEBページデータに含まれる前記単語群および前記単語群が形成するデータ構造を含むことを特徴とする請求項38に記載の記録媒体。

【請求項52】 前記分野辞書は、以前に前記検索ステ

ップの処理により用いられた前記第2の特徴語のリストに含まれる前記単語および前記単語群が形成するデータ構造を含むことを特徴とする請求項3に記載の記録媒体。

【請求項53】 前記重み修正ステップは、前記分野辞書に含まれる前記単語に対応する前記第1の特徴語の前記第1の重みを修正することを特徴とする請求項38に記載の記録媒体。

【請求項54】 既存の文書ファイルおよび所定の文書ファイルに関連する関連情報を提示する情報処理装置を制御するコンピュータに、

前記既存の文書情報から第1の特徴語を抽出するとともに、前記所定の文書情報から第2の特徴語を抽出する抽出ステップと、

前記抽出ステップの処理により抽出された、前記第1の特徴語の第1の重みを算出するとともに、前記第2の特徴語の第2の重みを算出する重み算出ステップと、
所定の条件に基づいて、前記重み算出ステップの処理により算出された前記第1の重みを修正する重み修正ステップと、

前記抽出ステップの処理により抽出された前記第1の特徴語、および前記重み修正ステップの処理により修正された前記第1の重みに基づいて、前記既存の文書情報に対応付けられた前記関連情報の取得を制御する取得制御ステップと、

前記第2の特徴語に基づいて、前記所定の文書情報に関連する前記既存の文書情報を検索する検索ステップと、
前記検索ステップの処理により検索された前記既存の文書情報に対応付けられている前記関連情報の表示を制御する表示制御ステップとを実行させるプログラム。

【請求項55】 前記取得制御ステップは、前記既存の文書情報における、複数の前記第1の特徴語の中から、前記第1の重みに基づいて重要語を選択し、前記重要語に関連する前記関連情報の取得を制御することを特徴とする請求項54に記載のプログラム。

【請求項56】 前記検索ステップは、前記第1の重みと前記第2の重みから類似度を算出し、前記類似度に基づいて、前記所定の文書情報に対応する前記既存の文書情報を検索することを特徴とする請求項54に記載のプログラム。

【請求項57】 前記既存の文書情報および前記所定の文書情報から属性情報を抽出する属性情報抽出ステップをさらに含み、

前記重み修正ステップは、前記属性情報抽出ステップの処理により抽出された前記属性情報に基づいて、前記第1の重みを修正することを特徴とする請求項54に記載のプログラム。

【請求項58】 前記属性情報抽出ステップの処理により抽出された前記属性情報に基づいて、1または複数の前記既存の文書情報をグループ化して文書情報群を作成

するグループ化ステップをさらに含むことを特徴とする請求項54に記載のプログラム。

【請求項59】 前記重み修正ステップは、特定の宛先への送信または前記特定の宛先からの受信の頻度に応じて、前記第1の重みを修正することを特徴とする請求項54に記載のプログラム。

【請求項60】 前記重み修正ステップは、前記第1の特徴語の種類に基づいて、前記第1の重みを修正することを特徴とする請求項54に記載のプログラム。

【請求項61】 前記既存の文書情報の前記第1の特徴語に関連付けて、前記取得制御ステップの処理により取得が制御された前記関連情報を用いてデータベースを構築するデータベース構築ステップをさらに含むことを特徴とする請求項54に記載のプログラム。

【請求項62】 所定のデータ構造で構成された単語群からなる、単語間の関係を示すデータベースである分野辞書を作成する分野辞書作成ステップをさらに含み、
前記重み修正ステップは、前記分野辞書作成ステップの処理により作成された前記分野辞書に基づいて、前記重み算出ステップの処理により算出された前記第1の重みを修正することを特徴とする請求項54に記載のプログラム。

【請求項63】 前記分野辞書は、前記単語群が、各前記単語の属する分野によってグループ化され、記録されているデータベースであることを特徴とする請求項54に記載のプログラム。

【請求項64】 前記分野辞書は、前記単語群が、前記単語間の関係に基づいて階層構造に配置され、記録されているデータベースであることを特徴とする請求項54に記載のプログラム。

【請求項65】 前記分野辞書は、前記単語群が、各前記単語に割り当てられた重要度に基づいてリスト化されたデータベースであることを特徴とする請求項54に記載のプログラム。

【請求項66】 前記分野辞書は、既存の辞書データベースに含まれる前記単語群および前記単語群が形成するデータ構造を含むことを特徴とする請求項54に記載のプログラム。

【請求項67】 前記分野辞書は、所定のWEBページデータに含まれる前記単語群および前記単語群が形成するデータ構造を含むことを特徴とする請求項54に記載のプログラム。

【請求項68】 前記分野辞書は、以前に前記検索ステップの処理により用いられた前記第2の特徴語のリストに含まれる前記単語群および前記単語群が形成するデータ構造を含むことを特徴とする請求項54に記載のプログラム。

【請求項69】 前記重み修正ステップは、前記分野辞書に含まれる前記単語に対応する前記第1の特徴語の前記第1の重みを修正することを特徴とする請求項54に

記載のプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、情報処理装置および方法、記録媒体、並びにプログラムに関し、特に、例えば、電子メールに含まれる情報の中から、ユーザの興味があると思われる単語および関連情報を取得してデータベースに蓄積し、その関連情報を効果的に表示させるようにした情報処理装置および方法、記録媒体、並びにプログラムに関する。

【0002】

【従来の技術】従来の、パーソナルコンピュータのデスクトップ（表示画面）に、いわゆるデスクトップマスコットと呼ばれるキャラクタを表示させるアプリケーションプログラムが存在する。

【0003】デスクトップマスコットは、例えば、電子メールの着信等をユーザに通知する機能やデスクトップ上を移動する機能などを有している。

【0004】ところで、例えば、ユーザが、電子メールとして送信する文書等を入力している時や、受信した文書を開覧している時などに、対象としている文書に関連する情報（以下、関連情報と記載する）を、デスクトップマスコットが提示するようになれば、ユーザによって利便性が向上するとともに、デスクトップマスコットに対して一層愛着を感じようになると考えられる。

【0005】そこで、例えば、特開2001-312515号公報には、電子メールなどの処理済みのテキストファイルを用いて自動的にデータベースを構築し、送受信などの対象とした文書に関連する関連情報をユーザに提示する方法が提案されている。

【0006】

【発明が解決しようとする課題】しかしながら、上記公報の発明においては、単に、対象となる文書に基づいて、データベースから関連情報を抽出して提示するだけであり、ユーザの興味の高さの反映されていなかった。

【0007】本発明はこのような状況に鑑みてなされたものであり、ユーザの興味の高さに応じて関連情報を提示することができるようにするものである。

【0008】

【課題を解決するための手段】本発明の情報処理装置は、既存の文書情報から第1の特徴語を抽出するとともに、所定の文書情報から第2の特徴語を抽出する抽出手段と、抽出手段により抽出された第1の特徴語の第1の重みを算出するとともに、第2の特徴語の第2の重みを算出する重み算出手段と、所定の条件に基づいて、重み算出手段により算出された第1の重みを修正する重み修正手段と、抽出手段により抽出された第1の特徴語、および重み修正手段により修正された第1の重みに基づいて、既存の文書情報に対応付ける関連情報を取得する

取得手段と、第2の特徴語に基づいて、所定の文書情報に関連する既存の文書情報を検索する検索手段と、検索手段により検索された既存の文書情報に対応付けられている関連情報の表示を制御する表示制御手段とを備えることを特徴とする。

【0009】取得手段は、既存の文書情報における、複数の第1の特徴語の中から、第1の重みに基づいて重要語を選択し、重要語に関連する関連情報を取得することができる。

【0010】検索手段は、第1の重みと第2の重みから類似度を算出し、類似度に基づいて、所定の文書情報に対応する既存の文書情報を検索することができる。

【0011】既存の文書情報および所定の文書情報から属性情報を抽出する属性情報抽出手段をさらに設けることができる。

【0012】属性情報抽出手段により抽出された属性情報に基づいて、1または複数の既存の文書情報をグループ化して文書情報群を作成するグループ化手段をさらに設け、抽出手段は、文書情報群から第1の特徴語を抽出し、重み算出手段は、第1の特徴語の第1の重みを算出し、取得手段は、第1の特徴語および第1の重みに基づいて、文書情報群に対応付ける関連情報を取得し、検索手段は、第2の特徴語に基づいて、所定の文書情報に対応する文書情報群を検索することができる。

【0013】重み修正手段は、属性情報抽出手段により抽出された属性情報に基づいて、第1の重みを修正することができる。

【0014】重み修正手段は、特定の宛先への送信または特定の宛先からの受信の頻度に応じて、第1の重みを修正することができる。

【0015】重み修正手段は、情報処理装置の使用者が作成した既存の文書情報から抽出された第1の特徴語の第1の重みを修正することができる。

【0016】重み修正手段は、他の既存の文書情報から抽出された、第1の特徴語と同じ特徴語の重みを、第1の重みと同様にさらに修正することができる。

【0017】重み修正手段は、第1の特徴語の種類に基づいて、第1の重みを修正することができる。

【0018】既存の文書情報の第1の特徴語に関連付けて、取得手段により取得された関連情報を用いてデータベースを構築するデータベース構築手段をさらに設け、検索手段は、第2の特徴語に基づいて、データベース構築手段により構築されたデータベースから、所定の文書情報に関連する既存の文書情報を検索することができる。

【0019】所定のイベントの発生を検知する検知手段をさらに設け、所定の文書情報は、検知手段により検知された所定のイベントの発生に対応する文書であるものとすることができる。

【0020】既存の文書情報および所定の文書情報は電

子メールであるものとすることができる。

【0021】所定のデータ構造で構成された単語群からなる、単語間の関係を示すデータベースである分野辞書を作成する分野辞書作成手段をさらに備え、重み修正手段は、分野辞書作成手段により作成された分野辞書に基づいて、重み算出手段により算出された第1の重みを修正することができる。

【0022】前記分野辞書は、単語群が、各単語の属する分野によってグループ化され、記録されているデータベースであるようにすることができる。

【0023】前記分野辞書は、単語群が、単語間の関係に基づいて階層構造に配置され、記録されているデータベースであるようにすることができる。

【0024】前記分野辞書は、単語群が、各単語に割り当てられた重要度に基づいてリスト化されたデータベースであるようにすることができる。

【0025】前記分野辞書は、既存の辞書データベースに含まれる単語群および単語群が形成するデータ構造を含むようにすることができる。

【0026】前記分野辞書は、所定のWEBページデータに含まれる単語群および単語群が形成するデータ構造を含むようにすることができる。

【0027】前記分野辞書は、以前に検索手段により用いられた第2の特徴語のリストに含まれる単語群および単語群が形成するデータ構造を含むようにすることができる。

【0028】前記重み修正手段は、分野辞書に含まれる単語に対応する第1の特徴語の第1の重みを修正するようすることができる。

【0029】本発明の情報記録方法は、既存の文書情報から第1の特徴語を抽出するとともに、所定の文書情報から第2の特徴語を抽出する抽出ステップと、抽出ステップの処理により抽出された、第1の特徴語の第1の重みを算出するとともに、第2の特徴語の第2の重みを算出する重み算出ステップと、所定の条件に基づいて、重み算出ステップの処理により算出された第1の重みを修正する重み修正ステップと、抽出ステップの処理により抽出された第1の特徴語、および重み修正ステップの処理により修正された第1の重みに基づいて、既存の文書情報に対応付ける関連情報の取得を制御する取得制御ステップと、第2の特徴語に基づいて、所定の文書情報に関連する既存の文書情報を検索する検索ステップと、検索ステップの処理により検索された既存の文書情報に対応付けられている関連情報の表示を制御する表示制御ステップとを含むことを特徴とする。

【0030】本発明の記録媒体に記録されているプログラムは、既存の文書情報から第1の特徴語を抽出するとともに、所定の文書情報から第2の特徴語を抽出する抽出ステップと、抽出ステップの処理により抽出された、第1の特徴語の第1の重みを算出するとともに、第2の

特徴語の第2の重みを算出する重み算出ステップと、所定の条件に基づいて、重み算出ステップの処理により算出された第1の重みを修正する重み修正ステップと、抽出ステップの処理により抽出された第1の特徴語、および重み修正ステップの処理により修正された第1の重みに基づいて、既存の文書情報に対応付ける関連情報の取得を制御する取得制御ステップと、第2の特徴語に基づいて、所定の文書情報に関連する既存の文書情報を検索する検索ステップと、検索ステップの処理により検索された既存の文書情報に対応付けられている関連情報の表示を制御する表示制御ステップとを含むことを特徴とする。

【0031】本発明のプログラムは、既存の文書情報から第1の特徴語を抽出するとともに、所定の文書情報から第2の特徴語を抽出する抽出ステップと、抽出ステップの処理により抽出された、第1の特徴語の第1の重みを算出するとともに、第2の特徴語の第2の重みを算出する重み算出ステップと、所定の条件に基づいて、重み算出ステップの処理により算出された第1の重みを修正する重み修正ステップと、抽出ステップの処理により抽出された第1の特徴語、および重み修正ステップの処理により修正された第1の重みに基づいて、既存の文書情報に対応付ける関連情報の取得を制御する取得制御ステップと、第2の特徴語に基づいて、所定の文書情報に関連する既存の文書情報を検索する検索ステップと、検索ステップの処理により検索された既存の文書情報に対応付けられている関連情報の表示を制御する表示制御ステップとをコンピュータに実行させる。

【0032】取得制御ステップは、既存の文書情報における、複数の第1の特徴語の中から、第1の重みに基づいて重要語を選択し、重要語に関連する関連情報の取得を制御することができる。

【0033】検索ステップは、第1の重みと第2の重みから類似度を算出し、類似度に基づいて、所定の文書情報に対応する既存の文書情報を検索することができる。

【0034】既存の文書情報および所定の文書情報から属性情報を抽出する属性情報抽出ステップをさらに含み、重み修正ステップは、属性情報抽出ステップの処理により抽出された属性情報に基づいて、第1の重みを修正することができる。

【0035】属性情報抽出ステップの処理により抽出された属性情報に基づいて、1または複数の既存の文書情報をグループ化して文書情報群を作成するグループ化ステップをさらに含むことができる。

【0036】重み修正ステップは、特定の優先度の送信または特定の優先から受信の頻度に応じて、第1の重みを修正することができる。

【0037】重み修正ステップは、第1の特徴語の種類に基づいて、第1の重みを修正することができる。

【0038】既存の文書情報の第1の特徴語に関連付け

で、取得制御ステップの処理により取得が制御された関連情報を用いてデータベースを構築するデータベース構築ステップをさらに含むことができる。

【0039】所定のデータ構造で構成された単語群からなる、単語間の関係を示すデータベースである分野辞書を作成する分野辞書作成ステップをさらに含み、重み修正ステップは、分野辞書作成ステップの処理により作成された分野辞書に基づいて、重み算出ステップの処理により算出された第1の重みを修正することができる。

【0040】前記分野辞書は、単語群が、各単語の属する分野によってグループ化され、記録されているデータベースであるようにすることができる。

【0041】前記分野辞書は、単語群が、単語間の関係に基づいて階層構造に配置され、記録されているデータベースであるようにすることができる。

【0042】前記分野辞書は、単語群が、各単語に割り当てられた重要度に基づいてリスト化されたデータベースであるようにすることができる。

【0043】前記分野辞書は、既存の辞書データベースに含まれる単語群および単語群が形成するデータ構造を含むようにすることができる。

【0044】前記分野辞書は、所定のWebページデータに含まれる単語群および単語群が形成するデータ構造を含むようにすることができる。

【0045】前記分野辞書は、以前に検索ステップの処理により用いられた第2の特徴語のリストに含まれる単語群および単語群が形成するデータ構造を含むようにすることができる。

【0046】前記重み修正ステップは、分野辞書に含まれる単語に対応する第1の特徴語の第1の重みを修正するようにすることができる。

【0047】本発明の情報処理装置および方法、並びにプログラムにおいては、既存の文書情報から第1の特徴語が抽出されるとともに、所定の文書情報から第2の特徴語が抽出され、第1の特徴語の第1の重みおよび第2の特徴語の第2の重みが算出され、所定の条件に基づいて算出された第1の重みが修正され、第1の特徴語および第1の重みに基づいて、既存の文書情報に対応付けられる関連情報が取得され、第2の特徴語に基づいて、所定の文書情報に関連する既存の文書情報が検索され、検索された既存の文書情報に対応付けられている関連情報が表示される。

【0048】

【発明の実施の形態】以下、図を参照して、本発明の実施の形態について説明する。

【0049】図1は、本発明を適用したデスクトップマシンの構成（以下、エージェントと記載する）をデスクトップ上に表示するアプリケーションプログラム（以下、エージェントプログラムと記載する）、電子メールを送受信するアプリケーションプログラム（以下、メーラ

（mailer）と記載する）、および、文書作成または編集するワードプロセッサプログラム3の構成を説明する図である。エージェントプログラム1乃至ワードプロセッサ3は、例えば、パーソナルコンピュータ（詳細は、図2を参照して後述する）にインストールされて実行されるものである。

【0050】エージェントプログラム1は、処理の対象となる文書の関連情報（後述）を蓄積してデータベースを構築する蓄積部11、処理の対象となる文書に対応する関連情報をユーザに提示する提示部12、および、エージェント72（図9）の表示等を制御するエージェント制御部13から構成される。

【0051】蓄積部11の文書取得部21は、メーラ2によって送受信した文書やワードプロセッサプログラム3によって編集された文書などのうち、自己が未処理の文書を取得して文書属性処理部22および文書内容処理部23に供給する。なお、以下においては、主に、メーラ2による電子メールとして送受信した文書を例として説明する。

【0052】文書属性処理部22は、文書取得部21から供給される文書の属性情報を抽出し、属性情報に基づき、文書をグループ化して、文書内容処理部23および文書特徴データベース作成部24に供給する。電子メールである場合、文書の属性情報としてヘッダに記述されている情報（対象となっているメールを特定するメッセージID、参照しているメールのメッセージID（References, In-Reply-To）、宛先（To, Cc, Bcc）、送信元（From）、日付（Date）、あるいは標題（Subject））などが抽出される。そして、抽出された属性情報に基づいて、1または複数の文書がグループ化される。以下、属性情報に基づいてグループ化された文書群（メールグループ）を「話題」と称する。

【0053】文書内容処理部23は、文書属性処理部22でグループ化された文書群（話題）の文面を抽出し、形態素解析を施して、単語（特徴語）に分類する。単語は、品詞（名詞、形容詞、動詞、副詞、接頭詞、感動詞、助詞、および助動詞）別に分類される。ただし、広範囲に亘って分布している単語、すなわち、例えば、ほぼ全ての文書に含まれている、「こんにちは」、「よろしく」あるいは「お願いします」等の名詞以外の品詞は、後述する、関連情報を検索するためのキーワードには成り得ないので、不要語として除外される。

【0054】文書内容処理部23は、不要語が除外された各単語の出現頻度および複数の文書に亘る分布状況を求め、グループ化された文書群（話題）毎に、各単語の重み（文書の主旨に関連する程度を示す値）を演算する。これにより、話題毎に、各単語の重みを構成要素とする特徴ベクトルが算出される。例えば、全話題の単語（特徴語）数の総計がn個の場合、すべての話題の特徴ベクトルは、n次元空間のベクトルとなり、次式（1）

で表わされる。

特徴ベクトル=(単語1の重み w_1 , 単語2の重み w_2 , ..., 単語 n の重み w_n) (1)

【0055】重みの演算には、例えば、tf-idf法(Salt on, G., Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer, Addison-Wesley, 1989)等が用いられる。この方法を用いると、例えば、話題Aに含まれる単語が n 個あった場合、話題Aの特徴ベクトルのうち、含まれる単語の該当する構成要素には重みが与えられ、話題Aに含まれない単語の構成要素は、頻度が0であるため、重みの値は0となる。

【0056】なお、本発明においては、話題毎に特徴ベクトルを算出するものとして説明するが、これに限らず、文書毎、または、その他の単位毎(例えば、所定期間(1週間)に蓄積された文書群毎)に特徴ベクトルを算出することも勿論可能である。

【0057】文書特徴データベース作成部24は、文書内容処理部22でグループ化された文書群(話題)毎の各文書の属性情報と、文書内容処理部23で算出された話題毎の特徴ベクトル(すなわち、話題内に含まれる全ての単語の重み)を時系列順にデータベース化してハードディスクドライブなどよりなる記憶部49(図2)に記録する。文書特徴データベース作成部24はまた、単語の重みに基づいて、所定の条件を満たす単語を選択し、その単語を、関連情報を検索するための検索用キーワード(重要語)として記録する。文書特徴データベース作成部24はさらに、検索用キーワードに関連情報検索部25に供給し、それに対応して関連情報検索部25から返される関連情報をも、検索用キーワードに対応付けて記録する。

【0058】関連情報検索部25は、文書特徴データベース作成部24が選択して記録した検索用キーワードに対する関連情報を検索し、検索結果のインデックスを文書特徴データベース作成部24に返す。キーワードの関連情報を検索するには、例えば、インターネット上の検索エンジンを用いる方法がある。この場合、WebページのURL(Uniform Resource Locator)やWebページのタイトルが、関連情報として文書特徴データベース作成部24に返される。

【0059】提示部12のイベント管理部31は、メールや電子メールの送信や受信を終えたことや、編集中の文書のテキストデータ量が所定の閾値を超えたことを検出して、データベース問い合わせ部32に通知する。以下、メール2が電子メールの送信や受信を終えたことや、編集中の文書のテキストデータ量が所定の閾値を超えたことをイベント発生と記載する。

【0060】イベント管理部31は、内部にタイマ31Aを内蔵しており、所定時間が経過したか否かを監視し、所定時間が経過した場合、その旨をデータベース問い合わせ部32に通知する。

【0061】データベース問い合わせ部32は、イベント管理部31からのイベント発生の通知に対応して、イベント発生に対応する文書(例えば、受信した電子メール)を取得し、文書内容処理部23の処理と同様に、その文書の形態素解析を施して、不要語を除外した単語を抽出し、各単語の重みを演算する。これにより、イベント発生に対応する文書の特徴ベクトルが算出される。

【0062】データベース問い合わせ部32はまた、文書特徴データベース作成部24が作成したデータベースを検索し、算出したイベント発生に対応する文書の特徴ベクトルと、データベースに記録されている話題毎の特徴ベクトルとの内積から類似度をそれぞれ算出する。データベース問い合わせ部32は、類似度が最も高い話題を抽出し、抽出した話題に含まれる各単語のうち、重みが所定の条件(詳細については図6を参照して後述する)を満たすものを選択し、選択された単語(重要語)に対応する関連情報を、イベント管理部31を介して、あるいは、直接、関連情報提示部33に供給する。

【0063】関連情報提示部33は、イベント管理部31を介して、あるいは、直接、データベース問い合わせ部32から供給される関連情報を表示部48(デスクトップ)上に表示させる。

【0064】なお、蓄積部11によるデータベースの更新は、所定のタイミングにおいて実行される。データベースの更新処理は、図28のフローチャートを参照して後述する。また、蓄積部11によるデータベースの更新時に、記憶部49に記録した特徴ベクトルが、例えば、メールのやりとりの回数、送受信の状態、または、品詞の種類(特定の地域や名称を示す固有名称)に応じて修正される。特徴ベクトルの修正処理は、図31、図34、および図36のフローチャートを参照して後述する。さらに、提示部12による関連情報の提示処理は、イベント管理部31がイベント発生を検知する時に実行される。

【0065】図2は、エージェンツプログラム1がインストールされて実行されるパーソナルコンピュータの構成例を示している。

【0066】パーソナルコンピュータは、CPU(Central Processing Unit)41を内蔵している。CPU41にはバス44を介して、入出力インターフェース45が接続されている。入出力インターフェース45には、キーボード、マウスなどの入力デバイスよりなる入力部46、処理結果としての例えば音声信号を出力する出力部47、処理結果としての画像を表示するディスプレイなどよりなる表示部48、プログラムや各種データを格納するハードディスクドライブなどよりなる記憶部49、LAN(Local Area Network)やインターネットを介してデータを通信するモデムなどよりなる通信部50、および、磁気

ディスク5、光ディスク3、光磁気ディスク4、または半導体メモリ5などのプログラム格納媒体に対してデータを読み書きするドライブ1が接続されている。バス44には、ROM (Read Only Memory) 42およびRAM (Random Access Memory) 43が接続されている。

【0067】本発明のエージェントプログラム1は、磁気ディスク2、光ディスク3、光磁気ディスク4、または半導体メモリ5に格納された状態でパーソナルコンピュータに供給され、ドライブ1によって読み出されて、記憶部49に内蔵されるハードディスクドライブにインストールされている。記憶部49にインストールされているエージェントプログラム1は、入力部46に入力されるユーザからのコマンドに対応するCPU41の指令によって、記憶部49からRAM43にロードされて実行される。なお、パーソナルコンピュータの起動時において自動的にエージェントプログラム1が実行されるように設定しておくようにしてもよい。

【0068】また記憶部49に内蔵されるハードディスクドライブは、エージェントプログラム1、メール2、およびワードプロセッサ3の他、WWW (World Wide Web) ブラウザなどのアプリケーションプログラムもインストールされており、エージェントプログラム1と同様に、入力部46に入力されるユーザからの起動コマンドに対応するCPU41の指令によって、記憶部49からRAM43にロードされて実行される。

【0069】次に、図3のフローチャートを参照して、エージェントプログラム1のデータベース作成処理について説明する。このデータベース作成処理は、エージェントプログラム1が実行する処理のうちの1つであり、エージェントプログラム1が実行された場合において、データベースが未だ作成されていないときに開始される。

【0070】ステップS1において、文書取得部21は、未処理の文書（例えば、エージェントプログラム1が実行される以前に送受信された電子メール）を、記憶部49に内蔵されるハードディスクドライブから取得して文書属性処理部22および文書内容処理部23に供給する。

【0071】ステップS2において、文書属性処理部22は、ステップS1の処理で文書取得部21により取得され供給された電子メールの文書から属性情報（メッセージID等のヘッダ情報）を抽出し、その属性情報に基づき、文書をグループ化して（すなわち、話題毎にグループ化して）、文書内容処理部23および文書特徴データベース作成部24に供給する。

【0072】ステップS3において、文書内容処理部23は、ステップS2の処理で文書属性処理部22によりグループ化された文書群（話題）の本文のテキストデータを抽出し、さらにテキストデータの形態素解析を施し、不要語を除く単語（特徴語）を抽出する。ステップ

S4において、文書内容処理部23は、ステップS3の処理で抽出された各単語の出現頻度、複数の文書に亘る分布状況等を求め、話題毎に、各単語の重みを、例えば、tf-idf法を用いて算出する。

【0073】ステップS5において、文書特徴データベース作成部24は、ステップS2の処理で文書属性処理部22によりグループ化された話題毎の各文書の属性情報と、ステップS3の処理で文書内容処理部23により算出されたその話題内に含まれる全ての単語の重み（特徴ベクトル）をデータベース化する。

【0074】具体的には、図4に示されるように、グループ化された各話題（話題A、話題B、話題C、・・・）を時系列順にソートした後、各話題にそれぞれ含まれる全単語1乃至nの重みを時系列順にソートして記憶部49に記録する。ここで、図4においては、例えば、重みA1は、話題Aにおける単語1の重みの値を示し、重みB2は、話題Bにおける単語2の重みの値を示している。さらに例えば、話題Bに単語1が含まれていない場合、重みB1は0となる。

【0075】図3の説明に戻る。ステップS6において、文書特徴データベース作成部24は、単語の重みが所定の閾値以上である単語を検索用キーワードに選択するか、あるいは、重みが大きい順にソートされ、予め指定される個数の単語を選択して、関連情報検索部25に供給する。関連情報検索部25は、文書特徴データベース作成部24からの検索用キーワード（重要語）を検索条件として、インターネット上の検索エンジンにアクセスして検索結果を取得し、得られたWebページのURLやタイトルを関連情報として文書特徴データベース作成部24に出力する。

【0076】ステップS7において、文書特徴データベース作成部24は、関連情報検索部25からの関連情報を検索用キーワードに対応付けて記憶部49に記録することにより、データベースを作成する。なお、ステップS6の処理とステップS7の処理は、ステップS5までの一連の処理に継続して実行する場合と、一連の処理に継続せず、所定のタイミングにおいて実行する場合がある。

【0077】以上のデータベース作成処理が実行されることにより、送受信した電子メールの文書に対応した関連情報がデータベース内に蓄積されることになる。なお、データベース作成処理は、エージェントプログラム1が実行された場合に開始されるものとしたが、任意のタイミングで開始させることも可能である。さらに、このようにして作成されたデータベースは、所定の条件が満たされたときに更新される（更新のタイミングについては、図29を参照して述べる）。

【0078】また、データベース作成処理をユーザが強制的に中断することができるようにするために、中断要求があった場合、中断された時点で処理済みの文書を記

録し、再開要求があった場合、未処理の文書から処理を再開するようにしてもよい。

【0079】次に、図5のフローチャートを参照して、エージェントプログラム1の関連情報提示処理について説明する。この関連情報提示処理は、図3を用いて上述したデータベース作成処理とは異なり、エージェントプログラム1が実行されている間、繰り返して実行される。

【0080】ステップS11において、エージェントプログラム1は、入力部46に入力されるユーザからのコマンドによって、エージェントプログラム1の終了が指示されたか否かを判定し、エージェントプログラム1の終了が指示されていないと判定した場合、ステップS12に進む。

【0081】ステップS12において、イベント管理部31は、イベント発生（メール2の電子メールの送受信の終了等）を監視し、イベント発生が検知されない場合、ステップS11に戻り、上述した処理が繰り返して実行される。

【0082】ステップS12において、イベント発生が検知された場合、ステップS13に進み、イベント管理部31は、イベント発生をデータベース問い合わせ部32に通知する。データベース問い合わせ部32は、イベント管理部31からのイベント発生の通知に対応して、イベント発生に対応する文書（送受信された電子メール）を取得し、その文書の形態素解析を施して、不要語を除外した単語（特徴語）を抽出し、各単語の重みを演算する。これにより、イベント発生に対応する文書の特徴ベクトルが算出される。

【0083】ステップS14において、データベース問い合わせ部32は、文書特徴データベース作成部24が作成したデータベースを探索し、ステップS13の処理で算出された特徴ベクトルと、データベースに登録されている話題毎の特徴ベクトルとの内積から類似度をそれぞれ算出し、類似度が所定の条件（例えば、類似度が最大、もしくは類似度が所定の閾値以上）を満たすものを抽出する。

【0084】また、より好ましくは、話題の選択に際しての条件として、過去一定期間に属する話題であって、最近送受信したメールに多く含まれている、または、集中して使われている単語を、特徴ベクトルの重みに関して上位後に持たない話題を選択するようにすると良い。話題を確定する日時としては、話題を構成する電子メールの初め、終わり、若しくは、平均の送信日時、または、話題を構成する電子メールの初め、終わり、若しくは、平均の受信日時等、日時を特定できれば何れでもよい。これにより、最近使われていたが過去にあった話題の内、現在の興味に類似するものを抽出することができ、ユーザに意外性を与えるとともに、ユーザの潜在的な興味を喚起することができる。

【0085】ステップS15において、データベース問い合わせ部32は、ステップS14の処理で抽出された話題に含まれる各単語のうち、重みの時系列推移に着目して、以下に示す条件1および条件2を満たす単語（重要語）を選択する。

【0086】なお、この処理は、必ずしも必須ではなく、ステップS6で取得され、ステップS14で選択された話題に関連付けられた関連情報の内、ステップS6での重みが一番大きなものから順に、または、今まで表示していない関連情報から優先的に表示するようにしても良い。また、表示する関連情報は1つに限られない。

【0087】ここで、図6を参照して、単語の選択条件について説明する。図6は、データベースに蓄積されている単語の重みの時系列推移の例を示している。

【0088】例えば、条件1は、「単語の重みが、現時点以前の所定の期間X（例えば、2週間）、所定の閾値A以下であること」である。また例えば、条件2は、「現時点以前の所定の期間Y（例えば、5週間）において、異なる2以上の話題で、単語の重みが所定の閾値B以上であること」である。なお、条件3「条件2における異なる2以上の話題のうち、最も古い話題と最も新しい話題が所定の期間Z以上離れていること」を追加すればさらに好ましい。

【0089】このように条件を用いることにより、ユーザの興味が高いと思われる単語（重要語）を選択することが可能となる。特に、条件1によって、現時点に近い話題に含まれる単語は除外されるので、ユーザが現時点で意識しており、意外性のないと思われる関連情報（新しすぎる情報）を改めて選択することができ、かつ、かなり以前の話題に含まれる単語も除外されるので、ユーザが現時点で思い出すことができないと思われる関連情報（古すぎる情報）を改めて選択することも避けることができる。

【0090】データベース問い合わせ部32は、このようにして選択した単語（重要語）に対応する関連情報を、イベント管理部31を介して、あるいは、直接、関連情報提示部33に供給する。

【0091】図5の説明に戻る。ステップS16において、エージェント制御部13は、ステップS15の処理で選択した単語が含まれている文書の属性情報を、選択（推薦）した理由として表示させるとともに、対応する関連情報を表示するか否かをユーザに問う入力ウィンドウ81（図14）をデスクトップに表示させる。

【0092】なお、話題は、グループ化された複数の文書からなるため、重要語が含まれる文書も複数存在する場合がある（すなわち、重要語が含まれている文書の属性情報が複数存在する場合がある）。そこで、例えば、重要語が含まれている文書のうち、最もまたは最新の文書の属性情報を表示させるようにするか、あるいは、任意に指定された文書の属性情報を表示させる。また、人

力ウィンドウ81を表示させずに、直接、デスクトップ上に関連情報を表示させるようにしてもよい。

【0093】ステップS17において、エージェントプログラム1は、入力部46に入力されるユーザからのコマンドによって、ステップS16の処理で表示された力ウィンドウ81に呼応して、ユーザが「見る」ボタンを選択したか否かを判定する。ステップS17において、ユーザが「見る」ボタンを選択したと判定された場合、ステップS18に進む。なお、入力ウィンドウ81には、「見る」ボタンおよび「見ない」ボタン以外にも他の情報を表示したり、あるいは、表示しないようにすることもできる。

【0094】ステップS18において、関連情報提示部33は、イベント管理部31を介してデータベース問い合わせ部32から供給された関連情報をデスクトップに表示させる。この関連情報は、1または複数個を提示することができる。

【0095】なお、関連情報として表示される情報は、キーワードが与えられた所定のデータベースに蓄積された情報であれば、Webページのタイトルでなくてもかまわない。例えば、所定のデータベースに蓄積されている情報のインデックスを表示するようにして、ユーザのアクセス指令に対応して、そのインデックスのさらに詳細な情報を表示できるようにもよい。

【0096】ステップS19において、エージェントプログラム1は、入力部46に入力されるユーザからのコマンドによって、ステップS18の処理により関連情報として表示されたWebページのタイトルに対して、ユーザがアクセスを指令したと判定した場合、ステップS20に進む。ステップS20において、Webブラウザが起動され、対応するWebページへのアクセスが行われる。

【0097】ステップS19において、ステップS18の処理により関連情報として表示されたWebページのタイトルに対して、ユーザが記録を指令したと判定された場合、ステップS21に進む。ステップS21において、エージェントプログラム1は、対応するWebページのタイトルおよびURLを、提示履歴を表示するスクラップ帳ウィンドウ74（図9）に記録する。なお、ステップS19において、ステップS18の処理により関連情報として表示されたWebページのタイトルに対して、ユーザから何の指令もなされずに所定の時間が経過したと判定された場合、ステップS20またはステップS21の処理はスキップされて、ステップS11に戻り、上述した処理が繰り返し実行される。

【0098】なお、ステップS17において、ユーザが「見る」ボタンを選択しないと判定された場合、ステップS18乃至ステップS21の処理はスキップされて、ステップS11に戻り、上述した処理が繰り返し実行される。さらに、ステップS11において、ユーザによりエージェントプログラム1の終了が指示されたらと判定さ

れた場合、処理は終了される。

【0099】ここで、関連情報提示処理に関して、イベント発生に対応する電子メールを効率的に取得する方法について説明する。

【0100】まず、大多数のメーラがメールの保持形式に関し、次のような4つの特徴を有していることに着目する。第1に、メーラにおける1つのフォルダは、パーソナルコンピュータにおける1つのメールボックスファイルに対応していること。第2に、新たに受信したメールは、特定のフォルダに格納されるようになっており、パーソナルコンピュータでは当該フォルダに対応するファイルの末尾に追加され、このとき、1つのファイルには一般に複数のメールが含まれるので、メールとメールとの境界に、特定の文字列パターン（メーラによって異なる）からなる行が挿入されていること。第3に、送信したメールの記録も、同様の形式で特定のファイルに保存されること。第4に、一般に、送受信したメールが含まれるファイルはサイズが大きいこと（数百KB乃至1MB）。

【0101】以上の第1乃至第4の特徴を考慮して、次の手順によってイベント発生に対応するメールを取得する。始めに、メールボックスファイルの更新日時を抽出し、新たなメールが追加されたか否かを判断する。次に、新たにメールが追加されたメールボックスファイルの末尾から先頭方向に1行ずつ操作して、メールとメールとの境界を示す特定の文字列を抽出する。境界を示す文字列を抽出した場合、その位置からメールボックスファイルの末尾までデータを抽出する。

【0102】このような手法を用いることにより、イベント発生に対応する電子メールを効率的に取得することが可能となる。

【0103】次に、関連情報提示処理に関して、同一のメールに対して何度も関連情報を提示しないようにする方法について説明する。まず、関連情報を提示したメールのメッセージIDを記録するためのデータ構造を設定する。そして、イベントが発生した場合、そのイベントに対応するメールのメッセージIDを取得して、設定したデータ構造と比較する。データ構造の中に同じメッセージIDがある場合、そのメールに対しては既に関連情報を提示しているので、関連情報を提示しない。一方、データ構造の中に同じメッセージIDがない場合、そのメールに対しては関連情報を提示していないので関連情報を提示し、メッセージIDをデータ構造に登録する。

【0104】このような手法を用いることにより、同一のメールに対して何度も関連情報を提示するような事態の発生を抑止することが可能となる。

【0105】次に、図7および図8のプロフローチャートを参照して、図4を用いて説明した関連情報提示処理のうち、主にエージェントの動作および台詞等について、詳細に説明する。

【0106】例えば、エージェントプログラム1が起動されている状態においてメーラ2が起動された場合、ステップS31において、エージェント制御部13は、例えば、図9に示されるように、メーラのウィンドウ（以下、メーラウィンドウと記載する）71の表示と重疊しない位置に、エージェント72を表示（登場）させる。

【0107】なお、エージェント72の登場は、例えば、図10A乃至図10Dに示す画像が順次表示されることによって、エージェント72が前転しながらデスクトップ上に出現する動画が表現される。エージェント72の登場とともに、エージェント72の台詞として吹き出し73と、保存されている関連情報が一覧表示されているスクラップ帳ウィンドウ74（後述）が表示される。吹き出し73の中には、例えば、図11に示されるように、登場の挨拶「おはよう、SAITOさん!」と、自己紹介「はく、あひびよ。」の台詞が表示される。

【0108】また、吹き出し73の表示と同期して、吹き出し73に表示された台詞の音声信号が音声合成部（図示せず）によって合成されて出力するようにすることができる。なお、吹き出し73に表示された台詞の音声信号は、吹き出し73に表示された言語（図16の例の場合、日本語）、あるいは、吹き出し73に表示された言語と同じ意味を持つ他の言語（例えば、英語の場合、「Good morning, SAITO.」「I'm Ahi!」）の音声信号が音声合成部によって合成されて出力されるようにしてもよい。そして、以降に表示される吹き出し73にも対応する音声信号が同期して出力される。

【0109】なお、吹き出し73や対応する音声は、出力しなくてもよい。以降、いずれの処理においても、吹き出し73および対応する音声は、エージェントプログラム1の規定により出力するか否かを設定するようにするか、あるいは、ユーザにより、出力するか否かを任意に設定できるようにする。

【0110】その後、エージェント72の表示は、ステップS32において、例えば、図12に示されるように、待機中の様子（手を俵に組み、つま先を上下させる）を示す動画に推移される。

【0111】ステップS33において、エージェントプログラム1は、入力部46に入力されるユーザからのコマンドに応じて、メーラ2が終了されたか否かを判定し、メーラ2が終了されていないと判定した場合、ステップS34に進む。

【0112】ステップS34において（上述した図5のステップS12に対応する）、メーラ2は、ユーザから何らかのコマンド（メールの送受信、メールの編集、あるいは関連情報の編集等）が入力されたか否かを判定し、何らかのコマンドが入力されたと判定した場合、ステップS35に進み、コマンドに対応する処理を開始する。

【0113】ステップS35において、エージェントプログラム1のイベント管理部31は、メールの送信、受信、または編集のコマンドが入力されたか否かを判定し、メールの送受信または編集のコマンドが入力されたと判定した場合、ステップS36に進む。

【0114】ステップS36において、エージェント制御部13は、エージェント72の表示を、図12に示した待機中の様子から、例えば、図13に示されるように、作業中の様子（手足を激しく移動する）を示す動画に推移させる。この間に、図5のステップS13乃至S15の処理（ユーザに推薦する関連情報を選択する処理）が実行される。

【0115】ステップS37において、エージェントプログラム1は、コマンドに対応して開始されたメーラ2の処理（例えば、メール送信など）が継続中であるか否かを判定し、メーラ2の作業中の処理が終了するまで判定処理を繰り返し実行する。すなわち、メーラ2の作業中の処理が終了するまで、エージェント制御部13は、エージェント72の表示を、図13に示した作業中の状態のまま待機させる。

【0116】ステップS37において、メーラ2の処理が継続中ではない、すなわち、コマンドに対応して開始されたメーラ2の作業中の処理が終了したと判定された場合、ステップS38に進む。

【0117】ステップS38において、エージェントプログラム1は、再度、入力部46に入力されるユーザからのコマンドに応じて、メーラ2が終了されたか否かを判定し、メーラ2が終了されていないと判定した場合、ステップS39に進む。

【0118】ステップS39（図5のステップS16に対応する）において、エージェント制御部13は、ステップS37のメーラ2の処理がメール送信であった場合、エージェント72の吹き出し73に、例えば、「今、Aさんにメール送ったけど、某月某日にAさんと（タイトル）について話していたよね。その中にできた（キーワード）について、関連するページを見つけたよ。見てみる?」と表示させる。

【0119】また例えば、ステップS37のメーラ2の処理がメール受信であった場合、台詞の前半が「今、Aさんからメールがきたけど」となり、ステップS37のメーラ2の処理がメールの編集であった場合、台詞の前半が「今、Aさんにメールを書いているけど」となる。

【0120】なお、「某月某日にAさんと（タイトル）」について話していたよね。」の部分は、関連情報が選択（推薦）された理由に相当するが、関連情報の選択理由の表示を、ステップS39において実行せずに、後述するステップS42の処理（関連情報の表示）の後に表示するようにしてもよい。また、関連情報の選択理由の表示をユーザの指示により任意のタイミング（例えば、メニューで理由をたずねるコマンドを用意するなど）で実

行するようにしてもよい。

【0121】また、タイムマシナによる一定時間経過時の提示(定期的提示)に関しては、「今、Aさんからメールがきたけど」など、特定のイベントを示すような表現ではなく、例えば、「某月某日にAさんと(タイトル)」について話していたよね。」の部分だけを表示するようにする。

【0122】さらに、これらの吹き出し73は、関連情報を提示する前に表示してもよい、あるいは、関連情報を提示した後に表示してもよい。

【0123】吹き出し73に隣接する位置には、例えば、図14に示されるように、入力ウィンドウ81が表示される。入力ウィンドウ81には、図15に示されるように、関連情報の表示を指示するときに選択する「見る」と示されたボタン、関連情報を表示させない時に選択する「見ない」と示されたボタン、関連情報が選択された背景(選択理由)の再表示を指示するときに選択する「背景をもう一度教えて」と示されたボタンが表示される。

【0124】入力ウィンドウ81が表示された状態で、ステップS40において、エージェント制御部13は、エージェント72の表示を、図14に示した待機中の様子を示す動画に推移させる。ステップS41(図5のステップS17に対応する)において、エージェントプログラム1は、入力ウィンドウ81の中の「見る」と示されたボタン、「見ない」と示されたボタン、または「背景をもう一度教えて」と示されたボタンのいずれがユーザにより選択されたか否かを判定する。なお、この入力ウィンドウ81は、表示しなくてもよい。

【0125】ステップS41において、入力ウィンドウ81の「見る」と示されたボタンが選択されたと判定された場合、ステップS42に進む。ステップS42(図5のステップS18に対応する)において、エージェント制御部13は、例えば、図16および図17に示されるように、関連情報として推薦URL91を表示させ、エージェント72の表示を、表示された推薦URL91を指し示す動画に推移させるとともに、吹き出し73に、台詞「どう?」を表示させる。推薦URL91には、通常、推薦されるWebページのタイトルが表示され、推薦URL91の上にマウスカーソルが置かれたときだけURLも重畳して表示される。推薦URL91は、マウスカーソルでドラッグすることにより移動可能である。

【0126】ステップS43(図5のステップS19に対応する)において、エージェントプログラム1は、表示した推薦URL91に対するユーザのコマンドを検出する。表示される推薦URL91に対するユーザのコマンドとしては、記録、アクセス、または消去等がある。

【0127】推薦URL91に対する記録コマンドは、例えば、記録する推薦URL91をスクラップ帳ウィンドウ74までドラッグアンドドロップする方法や、マウスで

右クリックし、表示されるメニューの中から記録を選択する方法などが考えられる。あるいは、推薦URL91は、すべて自動的に記録されるようにしてもよい。アクセスコマンドや消去コマンドについても同様に、Webブラウザのアイコンやゴミ箱のアイコンにドラッグアンドドロップする方法や、マウスで右クリックし、表示されるメニューの中から選択する方法、あるいはクリックパルにする方法などが考えられる。

【0128】ステップS43において、推薦URL91に対する記録コマンドが検出された場合、ステップS44(図5のステップS21に対応する)において、エージェント制御部13は、エージェント72の表示を、例えば、図18に示されるように、驚く動作に推移させる。スクラップ帳ウィンドウ74の中に、記録が指示された推薦URL91に対応するWebページのタイトルが追加される。

【0129】また、ステップS43で、推薦URL91に対するアクセスコマンドが検出された場合、ステップS44(図5のステップS20に対応する)において、エージェント制御部13は、エージェント72の表示を、例えば、図19Aおよび図19Bに示されるように、笑顔となって喜ぶ様子に推移させる。吹き出し73には、「わーい」と表示され、対応する音声信号が出力される。

【0130】また、ステップS43で、推薦URL91に対する消去コマンドが検出された場合、ステップS44において、エージェント制御部13は、エージェント72の表示を、例えば、図20Aおよび図20Bに示されるように、泣き顔となって悲しみ失望した様子に推移させる。吹き出し73には、「だめかあ」と表示され、対応する音声信号が出力される。

【0131】ステップS44の処理が終了した後、ステップS32に戻り、それ以降の処理が繰り返して実行される。

【0132】なお、ステップS41において、入力ウィンドウ81の「見ない」と示されたボタンが選択されたと判定された場合、ステップS32に戻り、それ以降の処理が繰り返して実行される。また、ステップS41において、入力ウィンドウ81の「背景をもう一度教えて」と示されたボタンが選択されたと判定された場合、ステップS39に戻り、ステップS39乃至S41の処理が繰り返される。

【0133】ステップS38において、メーラ2が終了されたと判定された場合、ステップS45に進み、エージェント制御部13は、吹き出し73に、「え、そんな」と表示させ、対応する音声信号を出力させた後、ステップS46において、エージェント72の表示を消失させる(図25を参照して後述する)。

【0134】ステップS35において、関連情報の編集を指示するコマンドが入力されたと判定された場合、ス

ステップS47に進み、関連情報提示部33は、関連情報編集用ウィンドウ（図示せず）を表示させ、エージェント制御部13は、エージェント72の表示を、図18に示した待機中の様子から、図17と同様に、関連情報編集用ウィンドウを指示する様子に推移させる。その後、ユーザが関連情報編集用ウィンドウに対して編集のための入力を開始すると、ステップS48において、エージェント制御部13は、エージェント72の表示を、関連情報編集用ウィンドウを指し示す様子から、図13に示した作業中の様子を示す動画像に推移させる。

【0135】ステップS49において、エージェントプログラム1は、関連情報編集処理が継続中であるか否かを判定し、関連情報編集処理が終了するまで判定処理を繰り返し実行する。すなわち、関連情報編集処理が終了するまで、エージェント制御部13は、エージェント72の表示を、図13に示した作業中の状態のまま待機する。

【0136】ステップS49において、関連情報編集処理が継続中でない、すなわち、コマンドに対して開始された関連情報編集処理が終了したと判定された場合、ステップS50に進む。

【0137】ステップS50において、エージェント制御部13は、エージェント72の表示を、図18と同様に、顔の様子に推移させる。吹き出し3には、「変更したよ」と表示され、対応する音声信号が出力される。その後、処理は、ステップS23に戻り、それ以降の処理が繰り返し実行される。

【0138】ステップS34において、メーラ2に対してユーザから何らかのコマンドが入力されない状態が所定の時間以上継続した場合、ステップS51に進む。ステップS51において、エージェント制御部13は、エージェント72の表示を、所定の時間が経過する毎に、移動の状態、遊びの状態、または睡眠の状態に順次推移させる。この待機中の処理の詳細について、図8のフローチャートを参照して説明する。なお、各処理は、エージェント制御部13が実行する。

【0139】ステップS61において、エージェント72の表示は、図12に示した待機中の状態から、例えば、図21または図22に示した画像を用いて表現される移動の状態に推移する。

【0140】エージェント72の移動は、表示されているウィンドウと重畳しないようにデスクトップ上を横方向あるいは縦方向に行われる。なお、アクティブであるウィンドウ（いまの場合、メーラウィンドウ71）を検出して、その周囲を横方向あるいは縦方向に行うようにしてもよい。エージェント72がデスクトップ上を横方向（例えば、右方向）に移動するときには、例えば、図21A乃至図21Dに示される画像が順次用いられることにより、いわゆるワープしたかのような動画表現が実現される。

【0141】具体的には、エージェント72の表示は、移動開始位置において、図21Aに示されるように、体の向きが移動する方向とされ、その後、指向している方向にジャンプすると、図21Bに示されるように、頭部から顔に消滅して行く。そして、移動終了位置において、図21Cに示されるように、脚部から顔に表示されて、最終的には、図21Dに示されるように全身が表示される。

【0142】エージェント72がデスクトップ上を上下方向に移動するときには、例えば、図22A乃至図22Gに示される画像が順次用いられる。すなわち、エージェント72は、移動開始位置において、図22Aに示されるように、自身の尻尾（先端がコンセントプラグの形状をしている）を手で握り、図22Bに示されるように、尻尾の先端を頭上付近に差し込む。

【0143】その後、エージェント72の表示は、図22C、図22Dに順次示されるように、体の下部から徐々にロープに変身し、図22Eに示されるように、1本のロープになってその状態で移動終了位置まで移動する。移動終了位置においては、図22F、図22Gに順次示されるように、頭部から顔に復元されて、最終的に全身が表示される。

【0144】このように、エージェント72の移動を、いわゆるワープによって表現したり、1本のロープに変身させて表現したりすることにより、移動中を表現するために使われるリソース（メモリなど）の量を軽減させることが可能となる。

【0145】図8の説明に戻る。ステップS62において、イベント（メールの送受信、メールの編集、あるいは関連情報の編集等を指示するコマンドの入力）が発生したか否かが判定される。イベントが発生していないと判定された場合、ステップS63に進む。

【0146】ステップS63において、エージェント72の表示が移動の状態に推移した後、所定の時間が経過したか否かが判定され、所定の時間が経過したと判定されるまで、ステップS62およびステップS63の処理が繰り返し実行される。ステップS63において、所定の時間が経過したと判定された場合、ステップS64に進む。

【0147】ステップS64において、エージェント72の表示は、移動の状態から、例えば、図23に示される画像で表現される「遊び」の状態に推移する。図23Aは、エージェント72が蛇を齧っている状態を示しており、図23Bは、エージェント72が尻尾の先端を上方に差し込み、そこを支点としてぶら下がり揺れている状態を示している。

【0148】ステップS65において、イベントが発生したか否かが判定される。イベントが発生していないと判定された場合、ステップS66に進む。ステップS66において、エージェント72の表示が遊びの状態に推

移した後、所定の時間が経過したか否かが判定され、所定の時間が経過したと判定されるまで、ステップS65およびステップS66の処理が繰り返し実行される。ステップS66において、所定の時間が経過したと判定された場合、ステップS67に進む。

【0149】ステップS67において、エージェント72の表示は、遊びの状態から、例えば、図24に示される画像で表現される「睡眠」の状態に推移する。ステップS68において、イベントが発生したか否かが判定され、イベントが発生するまで判定処理が繰り返し実行される。ステップS68において、イベントが発生したと判定された場合、図7のステップS35にリターンし、それ以降の処理が実行される。

【0150】なお、ステップS62、またはステップS65において、イベントが発生したと判定された場合も同様に、図7のステップS35にリターンし、それ以降の処理が実行される。

【0151】また、図8のフローチャートには図示していないが、待機中の処理が実行されている最中において、メーラ2が終了されたと判定された場合、ステップS46にリターンする。同様に、図7のステップS33において、メーラ2が終了されたと判定された場合、ステップS46に進む。

【0152】ステップS46において、エージェント制御部13は、エージェント72の表示を、例えば、図25Aおよび図25Bに示される画像で表現される「消失」の状態に推移させる。図25Aは、エージェント72が手を振りながら背を向けて遠方に立ち去る状態を示しており、図25Bは、エージェント72が徐々に小さくなってやがて消失される状態を示している。

【0153】なお、エージェント72の消去とともに、吹き出し73、スクラップ根ウィンドウ74、および推薦URL91等の表示も消去される。

【0154】以上のように、本発明によれば、電子メール等の文書から重みの高い単語（重要語）を抽出し、関連情報を推薦する一連の処理に対応して、エージェント72が動作するので、エージェント72に対して信頼性や親しみが感じられるようになる。

【0155】ところで、上述したエージェント72の動作および吹き出し73の中の台詞の表示、並びに、表示された台詞に対応する音声信号の出力については、本発明のエージェントプログラム1だけでなく、他のアプリケーション、例えば、ゲームやワードプロセッサのヘルプ画面等に適用することが可能である。さらに、テレビジョン受像機、ビデオカメラ、またはカーナビゲーション等のディスプレイに表示されるキャラクタに適用することも勿論可能である。

【0156】また、同一のパーソナルコンピュータを複数のユーザが操作する場合、エージェント72の種別を複数用意して、ユーザ毎に表示されるエージェント72

（図26）の種類を変えるようにしてもよい。また、エージェント72は、ユーザが好みのキャラクタを自由に作成し、編集できるようにしてもよい。さらに、同一のユーザが複数のパーソナルコンピュータ上でエージェントプログラム1を利用する場合、異なるパソコン上においても同じエージェント72が表示されるようにしてもよい。

【0157】なお、以上においては、エージェントプログラム1が実行されている場合、エージェント72は、常に登場しているものとして説明したが、推薦時にのみ表示させたり、その設定を変更することができる。

【0158】例えば、ユーザは、エージェントプログラム1が実行されている状態で、マウスををクリックし、図26に示されるようなメニューボックス101を表示させて、その中から、「いろいろな設定をする」の項目を選択することにより、図27に示されるような設定画面を表示させる。

【0159】図27の例の設定画面には、複数のタブが配置されており、「エージェント」と示されたタブがアクティブとされているとき、ユーザが選択または入力可能な、エージェントの名前、表示、効果音、推薦期間、推薦保存数、推薦するときのせりふ、および推薦データ更新などの項目が表示される。

【0160】ユーザは、これらの項目に対して、それぞれ、所望の情報（エージェントの名前）を入力したり、あるいは、所定の項目を選択することによって、自分好みにエージェント72および吹き出し73の表示状態、あるいは、推薦する関連情報の推薦期間時間や保存数などを設定することができる。

【0161】次に、蓄積部11によるデータベースの更新のタイミングについて説明する。データベースは、図3のフローチャートを用いて上述したデータベース作成処理によって作成されるが、次のような第1乃至第3の状況になった場合、データベースの更新が必要である。

【0162】すなわち、第1の状況として、データベースが作成または更新されてから所定の期間が経過している場合、データベース内の関連情報が古くなってしまうので更新が必要である。第2の状況として、データベースに蓄積されている関連情報のうちの所定の割合が提示済みとなった場合、データベース内の関連情報を繰り返し提示してしまったり、提示する関連情報が不足してしまうので更新が必要である。第3の状況として、特設抽出に用いた文書が、例えば電子メールである場合、電子メールの送受信を繰り返していると、その文書の内容が変化するので更新が必要である。

【0163】なお、データベースの更新が必要である状況になった場合（例えば、イベント管理部31がタイマ31Aを監視し、所定の期間が経過したとき）、ユーザに対して更新を指示するように促すこともできし、ユーザに対する促しを実行せずに、自動的にデータベー

を実行するように設定することも可能である。また、ユーザが指定する任意のタイミングで更新することも勿論可能である。

【0164】前記第1乃至第3の状況を考慮したデータベース更新処理について、図28のフローチャートを参照して説明する。このデータベース更新処理は、エージェントプログラム1が実行する処理のうちの1つであり、エージェントプログラム1の起動とともに開始され、エージェントプログラム1が終了されるまで繰り返して実行される。なお、この処理を開始するにあたって、図3を用いて説明したデータベース作成処理が実行されており、データベースが既に作成されているものとする。

【0165】ステップS81において、エージェントプログラム1の蓄積部11は、作成済みのデータベースの更新が必要であるかを判定し、更新が必要であると判定されるまで待機する。この判断基準は、例えば、図29に示すようなユーザインタフェースの画面を用いて予めユーザが設定するものとする。図29の例では、4つの条件が示されており、左側の□(チェックボックス)がユーザによってチェックされた場合、対応する条件が有効となる。なお、1番目の条件では回数が設定可能とされており、3番目の条件では日数が設定可能とされている。

【0166】ステップS81において、更新が必要であると判定された場合、ステップS82に進み、さらに蓄積部11は、データベースを自動的に更新するように設定されているか否かを判定し、自動的に更新するように設定されていないと判定した場合、ステップS83に進む。一方、ステップS82において、自動的に更新するように設定されていると判定された場合、ステップS83の処理はスキップされる。

【0167】ステップS83において、エージェントプログラム1の提示部12は、データベースの更新が必要である旨をユーザに通知するとともに、さらに、その通知に対応して、ユーザから更新の指示がなされたか否かを判定する。ユーザから更新の指示がなされたらと判定された場合、ステップS84に進み、ユーザから更新の指示がなされないと判定された場合、ステップS81に戻り、上述した処理が繰り返して実行される。

【0168】ステップS84において、エージェントプログラム1の蓄積部11は、データベースを更新する。具体的には、文書取得部21乃至文書内容処理部23が、電子メールのメールボックスファイル(例えば、mbox等の拡張子が付加されていることが多い)を検出し、その更新日時を取得して、以前に取得した更新日時と比較して、異なる目付と異なるファイルサイズであれば、ファイルが更新されていると判断し、追加または変更された部分を抽出する。この場合、メールのグループ化、ヘッダの解析、形態素解析、特徴ベクトル算出等、一連

のファイル内の分析が行われ、得られる重要語が関連情報検索部25に供給される。

【0169】ただし、メールグループ(話題)が変化せず(所定の話題に追加されるメールがなく)、分析の結果、更新以前の重要語(検索用キーワード)と更新後の重要語が同じであれば、重み等の計算値だけを変更し、関連情報検索部25による関連情報の検索を実行しないようにしてもよい。また、関連情報をインターネット等のネットワーク上で検索する場合、ネットワークに接続されているか否かが検出され、ネットワークに接続されていない場合、関連情報の検索を行わず、その後、ネットワークに接続されたときに関連情報を検索するか否かをユーザに問うようにしてもよい。

【0170】同じ関連情報を何度も推薦(提示)しないようにするために、あるメールグループの関連情報を、所定の回数以上推薦したら更新が必要と判断する」との条件に関連して、取得したメールと類似性の高いメールグループ(話題)を選択する際に、同じメールグループから何度も推薦を行わないように、次のような処理を行う。

【0171】メールグループ自体に推薦の優先度の順位を付与し(例えば、メールグループ内での特徴語の重みの最大値をそのメールグループの優先度の値とし、優先度の値を降順に並べたものを優先度の順位として付与する)、一度推薦を行ったメールグループを優先順位の最後尾に並び替えるようにする。このようにすることによって、類似度の範囲内にあるメールグループでも、同じメールグループから推薦する頻度が減少する。また、優先順位の変更だけなので、関連情報を大量に検索して準備しておけば、なるべく同じメールグループからの推薦が渡り、かつ、情報自体も不足することなく用いることができる。

【0172】これに関連して、特徴抽出に用いる話題内の文書量に応じて、類似する話題を抽出する際の範囲を変化させることができるようにする。具体的には、特徴抽出する話題の文書量またはデータサイズに応じて何段階かの類似度の範囲を設定する。例えば、ある話題に含まれる文書量が10ファイル以内である場合は類似度を0.01以上、11ファイル以上50ファイル未満の場合は類似度を0.03以上、50ファイル以上である場合は類似度を0.05以上とする。または、ある話題の文書の容量が500KB未満である場合は類似度を0.01以上、500KB以上である場合は類似度を0.02以上とする。

【0173】そして、予め設定された類似度の範囲のうち、優先度の高い話題から検索された関連情報を提示するようにする。このようにすると、文書量の減少により、データベースの内容が更新されると、類似度の範囲が変化し、類似度の範囲が狭くて関連情報が不足したり、反対に、類似度の範囲が広すぎてユーザにとってあ

より関連性が明確でない関連情報が提示されるような事態の発生を抑制することが可能となる。

【0174】以上説明したように、データベース更新処理においては、追加された文書や変更された文書だけを処理の対象とするので、データベース作成処理を繰り返して実行する場合に比べて、処理時間が短縮される。

【0175】また、データベース更新処理において、ユーザの興味をより反映させることができるようにするために、算出された単語の重みを修正することができる。次に、単語の重みの修正処理について、いくつか例を挙げて説明するが、その前に、図30を参照して、データベースに既に記録されている、各話題の特徴ベクトルの例について説明する。

【0176】図30に示されるように、話題Aの特徴ベクトル(単語aの重み、単語bの重み、単語cの重み、単語dの重み、単語eの重み、単語fの重み、単語gの重み)は、(0.54, 0.02, 0.0, 0.01, 0)であり、話題Bの特徴ベクトルは、(0.01, 0.41, 0.33, 0.16, 0.64)であり、話題Cの特徴ベクトルは、(0.51, 0.35, 0.0, 0.48, 0.37, 0)である。

【0177】以下、図30に示した各話題(話題A、話題B、および話題C)に含まれる単語の重みを修正する場合の処理について、第1乃至第3の例を挙げて順に説明する。

【0178】まず、第1の例の重み修正処理について、図31のフローチャートを参照して説明する。

【0179】ステップS101、S102において、文書特徴データベース作成部24は、メールの相手毎に、返信を含む返信した回数(やりとりした頻度)をカウントし、返信の頻度の多い順に相手をソートする。なお、返信(返信)した回数をカウントするだけでなく、受信した回数もカウントするようにしてもよい。

【0180】例えば、メールのやりとりをしている相手が、Aさん、Bさん、Cさん、およびDさんであり、やりとりの頻度が、Aさんとは3回、Bさんとは2回、Cさんとは28回、Dさんとは4回であるとカウントされた場合、図32に示されるように、やりとりした頻度が多い順に、Aさん、Cさん、Dさん、Bさんにソートされる。

【0181】ステップS103において、文書特徴データベース作成部24は、ステップS102の処理でソートされた、やりとりの頻度が上位のメールの相手を抽出する。

【0182】具体的には、例えば、予め設定されている上位n(例えば、n=2)番目までのメールの相手を抽出するか、所定の頻度m(例えば、m=10)回以上のメールの相手を抽出するか、上位n番目以上、かつ、所定の頻度m回以上のメールの相手を抽出するか、あるいは、その他の方法によってメールの相手を抽出する。ここでは、メールの相手の抽出方法は特に問わない。

【0183】例えば、上位2番目までのメールの相手を抽出する場合、図32に示した4人のメールの相手の中から、AさんとCさんが抽出される。

【0184】ステップS104において、文書特徴データベース作成部24は、ステップS103の処理で抽出された相手とのメール(話題)に含まれる単語の重みを修正する。

【0185】具体的には、AさんとCさんとのメールのやりとりが、例えば、話題Aであった場合、その話題Aに含まれる単語は、図30に示されるように、b、c、fである(単語a、d、e、gの重みがそれぞれ0である)ことは、それらの単語は、話題Aには含まれていない。そして、話題Aに含まれる単語b、c、fの重みがそれぞれ修正される。

【0186】より具体的には、単語b、c、fの重みに、予め設定されている重み係数k(例えば、1.2)を乗算して修正する(相対的に大きくする)。すなわち、話題Aに含まれる、単語bの重み(=0.54)、単語cの重み(=0.02)、単語fの重み(=0.01)をそれぞれ1.2倍する。これにより、図33Aに示されるように、話題Aの特徴ベクトルが、(0.65, 0.02, 0.12, 0.02, 0.02, 0.01, 0)に修正される。

【0187】以上のようにして、ユーザがより興味を感じている話題に含まれる単語の重みを修正することができる。

【0188】また例えば、AさんとCさんとの話題Aに含まれる単語b、c、fが、AさんCさん以外の人(BさんとDさん)との話題に含まれる場合にも、該当する単語の重みに、同様の重み係数k(いまの場合、1.2)を乗算して修正することもできる。すなわち、図30に示した話題Bには単語cが含まれており、話題Cには単語b、fが含まれており、それらの重みをそれぞれ1.2倍する。これにより、図33Bに示されるように、話題Bの特徴ベクトルが、(0.01, 0.49, 0.41, 0.33, 0.33, 0.16, 0.64)に修正されたとともに、話題Cの特徴ベクトルが、(0.51, 0.42, 0.35, 0.16, 0.48, 0.44, 0.37, 0.12)に修正される。

【0189】以上のように、メールのやりとりの多い相手との話題に含まれる単語が、他の話題にも含まれている場合、その単語の重みも修正することによって、メールのやりとりの少ない相手との話題の中から、ユーザがより興味を持ちやすい単語を重要語として選択することができる。

【0190】さらにまた、メールのやりとりの頻度または順位によって、重み係数kの値を変更するようにしてもよい。具体的には、Aさん(頻度1位)とのメールのやりとりが、例えば、話題Bであり、その話題Bに含まれる単語a、c、d、e、gには、重み係数k=1.5

を乗算し、Cさん(頻度2位)とのメールのやりとりが、例えば、話題Cであり、その話題Cに含まれる単語a, b, c, fには、重み係数 $k=1.2$ を乗算する。これにより、図33Cに示されるように、話題Bの特徴ベクトルが、 $\{0.02 (=0.01 \times 1.5), 0, 0.62 (=0.41 \times 1.5), 0.50 (=0.33 \times 1.5), 0.24 (=0.16 \times 1.5), 0, 0.96 (=0.64 \times 1.5)\}$ に修正されるとともに、話題Cの特徴ベクトルが、 $\{0.61 (=0.51 \times 1.2), 0.42 (=0.35 \times 1.2), 0, 0, 0.58 (=0.48 \times 1.2), 0.44 (=0.37 \times 1.2), 0\}$ に修正される。

【0191】以上のようにして、ユーザがより興味を感じている話題に含まれる単語の重みを、メールの相手とのやりとりの頻度に応じて修正することができる。また、重み係数 k は、頻度に比例した値でも、任意に設定した値でも、いずれでもよい。

【0192】次に、第2の例の重み修正処理について、図34のフローチャートを参照して説明する。

【0193】ステップS111において、文書特徴データベース作成部24は、話題毎にグループ化されている複数のメールの中から、送信(返信を含む)メールを特定し、その送信メールに含まれる単語を抽出する。ステップS112において、文書特徴データベース作成部24は、ステップS111の処理で抽出された返信メールに対応する元の受信メールを特定し、その受信メールに含まれる単語を抽出する。

【0194】これにより、例えば、話題Aに対して送信したメールに含まれる単語b, cが抽出され、話題Bに対して返信したメールに含まれる単語e, gが抽出され、その返信メールに対応する元の受信メールに含まれる単語cが抽出される。

【0195】ステップS113において、文書特徴データベース作成部24は、ステップS111およびS112の処理でそれぞれ抽出された単語の重みを修正する。

【0196】具体的には、送信(返信を含む)メールに含まれる単語の重みに、予め設定されている重み係数 k_1 (例えば、1.5)を乗算して修正するとともに、返信メールに対応する元の受信メールに含まれる単語の重みに、重み係数 k_2 (例えば、1.2)を乗算して修正する。これにより、図35に示されるように、話題Aの特徴ベクトルが、 $\{0, 0.81 (=0.54 \times 1.5), 0.03 (=0.02 \times 1.5), 0, 0, 0.01, 0\}$ に修正され、話題Bの特徴ベクトルが、 $\{0.01, 0, 0.49 (=0.41 \times 1.2), 0.33, 0.24 (=0.16 \times 1.5), 0, 0.96 (=0.64 \times 1.5)\}$ に修正される。

【0197】以上のように、ユーザが送信または返信したメールに含まれる単語の重みを修正することによって(ユーザが積極的に使用する単語の重みを修正することによって)、ユーザがより興味を感じている単語を重要語として選択することができる。なお、受信のみのメールについては、興味がないものとして重みを修正しない

ようにする。

【0198】また、送信、返信、または受信に応じて、それぞれ重み係数を設定するようにしたり、あるいは、特定の話題において、送信(返信を含む)メールに含まれる単語および返信メールに対応する元の受信メールに含まれる単語が、他の話題にも含まれている場合、その単語の重みを修正するようにしてもよい。

【0199】次に、第3の例の重み修正処理について、図36のフローチャートを参照して説明する。

【0200】ステップS121において、文書特徴データベース作成部24は、話題内の各単語のうち、特定の地域や名称を示す固有名詞を抽出する。これにより、例えば、単語b, fが固有名詞として抽出される。

【0201】ステップS122において、文書特徴データベース作成部24は、ステップS121の処理で抽出された単語の重みを修正する。

【0202】具体的には、話題に含まれる、固有名詞である単語の重みに、予め設定されている重み係数 k (例えば、1.5)を乗算して修正する。すなわち、図30に示した各話題に含まれる単語b, fの重みをそれぞれ1.5倍する。これにより、図37に示されるように、話題Aの特徴ベクトルが、 $\{0, 0.81 (=0.54 \times 1.5), 0.02, 0, 0, 0.02 (=0.01 \times 1.5), 0\}$ に修正され、話題Cの特徴ベクトルが、 $\{0.51, 0.53 (=0.35 \times 1.5), 0, 0, 0.48, 0.56 (=0.37 \times 1.5), 0\}$ に修正される(話題Bには、単語b, fは含まれていない)。

【0203】以上のようにして、ユーザがより興味を感じている単語の重みを修正することができる。また、固有名詞の種類(例えば、地名、組織名、商品名、または人名など)に応じて、重み係数 k の値を変更するようにしてもよい。ただし、人名(単語)の重みを修正したとしても、それを重要語として選択することはできないので、人名の重みは修正しないようにする。

【0204】また、メールのやりとり頻度、送受信の状態、または品詞(固有名詞)によって、単語の重みを修正する場合を説明したが、勿論、これらを組み合わせて、単語の重みを修正することもできる。さらに、ユーザの興味により反映されるものであれば、単語の重みの修正処理は、これに限られるものではない。

【0205】例えば、単語が属する分野ごとに単語をまとめた分類した分野辞書を設け、この分野辞書に含まれるか否かによって、文書特徴データベース作成部24が、データベースに含まれる単語の重みを修正するようにしてもよい。

【0206】図38は、本発明を適用したエージェントプログラムの機能ブロックの、他の構成例を示す図である。

【0207】図38において、分野辞書201は、例えば、人文、芸術、または科学等の領域、分野、またはカテゴリー毎に対応する単語をまとめたデータベースであ

る。分野辞書201には、ユーザが興味を示す分野を中心とする複数の分野が形成されており、各分野には、対応する単語が含まれている。

【0208】文書特徴データベース作成部24は、この分野辞書201を参照し、作成したデータベースに含まれる単語がこの分野辞書201に登録されているか否かを判定し、目的の単語が分野辞書201に含まれている場合、その単語の重みを修正する。

【0209】すなわち、データベースに含まれる単語が分野辞書201に含まれている場合(ユーザの興味がある分野の単語である場合)、文書特徴データベース作成部24は、その単語の重みを修正する。これにより、データベースの単語の重みづけにユーザの興味を反映することができる。

【0210】次に、図39のフローチャートを参照して、文書特徴データベース作成部24による分野辞書を用いた重み修正処理について説明する。

【0211】最初に、ステップS141において、文書特徴データベース作成部24は、分野辞書201を参照し、データベースに登録されている単語に対応する単語を検索する。文書特徴データベース作成部24は、分野辞書201より、データベースに登録されている単語と同一の単語、同じ分野に属する単語等のように、対応する単語を検索する。

【0212】そして、ステップS142において、文書特徴データベース作成部24は、検索結果に基づいて、対応する単語が存在するか否かを判定し、存在すると判定した場合、ステップS143に処理を進める。

【0213】ステップS143に処理を進めた文書特徴データベース作成部24は、分野辞書201より検索された単語を抽出し、ステップS144において、データベースを参照して、抽出した単語に対応する重みを修正し、分野辞書201を用いた重み修正処理を終了する。

【0214】また、ステップS142において、検索結果より、データベースに含まれる単語に対応する単語が存在しないと判定した場合、文書特徴データベース作成部24は、ステップS143および144の処理を省略して、分野辞書201を用いた重み修正処理を終了する。

【0215】具体的には、例えば、ある話題の特徴ベクトル(単語Aの重み、単語Bの重み、単語Cの重み、単語Dの重み、単語Eの重み、単語Fの重み、単語Gの重み)が(1.1, 0.1, 0, 0, 1.4, 5.1, 3.8)である場合、データベース問い合わせ部32により2語が最重要語として選択されるとすると、最重要語は、単語Fおよび単語Gとなる。

【0216】これに対して、分野辞書201に、単語A、単語B、単語D、単語H、単語K、および単語Lが含まれており、文書特徴データベース作成部24が、データベースに登録されているこれらの単語の重みをう

に修正したとすると、上述した特徴ベクトルは、(5.1, 5.1, 0, 0, 1.4, 5.1, 3.8)に修正されており、データベース問い合わせ部32は、単語Aおよび単語Fを最重要語として選択する。

【0217】なお、以上においては、分野辞書201に対応する単語が含まれる、データベースの単語の重みを全て一定の割合で修正するように説明したが、これに限らず、分野辞書201に含まれる単語毎に修正する割合を変化させるようにしてもよい。

【0218】以上のようにして、文書特徴データベース作成部24は、分野辞書201を用いて重み修正処理を行う。

【0219】以上においては、文書特徴データベース作成部24が分野辞書201に含まれる単語に対応する、データベースに含まれる単語の重みを修正するように説明したが、これに限らず、分野辞書201に含まれる単語に対応する、データベースの単語を、最重要語に指定するようにしてもよい。

【0220】次に、文書特徴データベース作成部24による分野辞書を用いた重み修正処理の他の例について、図40のフローチャートを参照して説明する。

【0221】文書特徴データベース作成部24は、最初に、ステップS161において、分野辞書201を参照し、データベースに登録されている単語に対応する単語を検索する。文書特徴データベース作成部24は、分野辞書201より、データベースに登録されている単語と同一の単語、同じ分野に属する単語等のように、対応する単語を検索する。

【0222】そして、ステップS162において、文書特徴データベース作成部24は、検索結果に基づいて、対応する単語が存在するか否かを判定し、存在すると判定した場合、ステップS163に処理を進める。

【0223】ステップS163に処理を進めた文書特徴データベース作成部24は、分野辞書201より検索された単語を抽出し、ステップS164において、データベースを参照して、抽出した単語を最重要語としてデータベースに登録し、分野辞書201を用いた重み修正処理を終了する。

【0224】また、ステップS162において、検索結果より、データベースに含まれる単語に対応する単語が存在しないと判定した場合、文書特徴データベース作成部24は、ステップS163および164の処理を省略して、分野辞書201を用いた重み修正処理を終了する。

【0225】具体的には、例えば、ある話題の特徴ベクトル(単語Aの重み、単語Bの重み、単語Cの重み、単語Dの重み、単語Eの重み、単語Fの重み、単語Gの重み)が(1.1, 0.1, 0, 0, 1.4, 5.1, 3.8)である場合、データベース問い合わせ部32により2語が最重要語として選択されるとすると、最重要語は、単語Fおよび

び単語Gとなる。

【0226】これに対して、分野辞書201に、単語A、単語B、単語D、単語H、単語K、および単語Iが含まれており、文書特徴データベース作成部24が、データベースに記録されているこれらの単語を最重要語としてデータベースに登録したとすると、上述した特徴ベクトルは、(1.1, 0.1, 0, 0, 1.4, 5.1, 3.8)のままだが、データベース問い合わせ部32は、最重要語に指定された単語Aおよび単語Bを最重要語として選択する。なお、単語Dも分野辞書201に含まれているが、単語Dの重みの値が「0」であるため、指定されていない。

【0227】以上のようにして、文書特徴データベース作成部24は、分野辞書201に含まれる単語に対応する、データベースに登録されている単語を最重要語として指定しており、データベース問い合わせ部32は、最重要語に指定された単語を最重要語として選択する。

【0228】以上において、分野辞書201に含まれる単語は6個として説明したが、これに限らず、いくつであってもよい。また、その分野辞書201の構造はどのような構造であってもよい。

【0229】次に、以上に説明したような分野辞書201の作成方法について説明する。

【0230】例えば、分野辞書201は、一般的に市販または配布されているような、階層構造またはグループ構造などで形成されたシソーラス辞書やオントロジー辞書を用いて作成される。

【0231】これらの辞書は、例えば、予め本体アプリケーションと同じ配布メディアに同梱され、CPU41により、ドライブ51に装着されたそのメディアから記憶部49にインストールされる。そして、CPU41は、インストールされた辞書を参照し、分野辞書201を生成する。

【0232】すなわち、CPU41は、インストールされた辞書の全部または一部より単語を抽出し、さらに、その辞書の構造(階層、グループ等)を反映させながら、分野辞書201を生成する。

【0233】分野辞書201の更新する場合、CPU41は、生成する場合と同様に、ドライブ51に装着されたメディアから新たにインストールされた辞書を参照して、更新処理を行う。

【0234】また、参照される辞書は、メール等により、インターネットに代表されるネットワークから取得されるようにしてもよい。このとき、CPU41は、上述した場合と同様に分野辞書201を生成する。また、この場合、CPU41は、所定の間隔で定期的に、またはユーザの指示に基づいて、ネットワークより新たな辞書を取得し、分野辞書201の情報を更新する。

【0235】さらに、参照される辞書は、上述したシソーラス辞書やオントロジー辞書以外にも、例えば、感度

良く頻繁に抽出する単語をリスト化したものであってもよい。また、ユーザの入力により直接分野辞書201が作成されるようにしてもよい。

【0236】また、以上に説明した分野辞書201は、上述したような既存の辞書以外から作成されるようにしてもよい。例えば、分野辞書201は、オンラインショッピングのサイト、掲示板サイト、または、カテゴリ別検索サイト等のようなインターネット上のWEBページを用いて作成される。

【0237】CPU41は、通信部50を介して、接続されているインターネット上から上述したようなWEBページを取得する。そして、そのWEBページに基づいて、分野辞書201を生成する。すなわち、CPU41は、取得したWEBページに含まれる、例えば商品名等のような単語を抽出し、さらにそのWEBページの構造(階層、グループ等)を反映させながら、分野辞書201を生成する。

【0238】この場合、CPU41は、所定の間隔で定期的に、またはユーザの指示に基づいて、ネットワークより新たなWEBページを取得し、分野辞書201の情報を更新する。

【0239】また、例えば、関連情報検索部25が図3のステップS6において、文書特徴データベース作成部24からの検索用キーワード(重要語)を検索条件として、インターネット上の検索エンジンにアクセスして検索結果を取得する際に、外部のサーバ(図示せず)が使用された検索用キーワードを取得し、そのキーワードに基づいて情報を作成するようにし、その作成された情報に基づいて、分野辞書201が作成されるようにしてもよい。

【0240】外部に設けられたサーバ(図示せず)は、関連情報検索部25が用いた検索用キーワードを取得して蓄積し、使用頻度に応じてランキングされた単語のリストを生成する。

【0241】そして、CPU41は、所定のタイミングでそのリストを取得し、そのリストに基づいて、分野辞書201を生成する。

【0242】例えば、図41に示されるように、サーバは、取得した検索用キーワードより、ユーザ(Xさん、Yさん、)毎のリストを生成し、各単語が使用された頻度に応じて、単語をグループ化する。

【0243】図41において、サーバは、検索される頻度の高い単語を単純にリスト化しただけの辞書ではなく、同じトピック(TopicX01乃至TopicX04、並びに、TopicY01乃至TopicY03)の中から用いられる検索用キーワードの組み合わせに基づいて、検索用キーワードをグループ化(6p1, 6p2)したリストを生成している。

【0244】また、それ以外にも、例えば、同じような検索用キーワードを用いているユーザを、検索用キーワードを成分とするベクトル空間の内積をとり、一定範囲

内のユーザ同士をまとめるなどして、グループ化し、同じグループ内のユーザの用いる検索用キーワードだけを抽出するようにしてもよい。

【0245】さらに、上述した以外にも、例えば、PW41が、関連情報にアクセスする操作や関連情報を保存するような操作についての操作情報を取得して、記憶部49等に記録し、所定のタイミングでその検索情報をサーバに供給するようにし、サーバは、その操作情報に基づいて、ユーザの評価が高い(ユーザが対応する関連情報に興味があった)検索用キーワードを優先的にリスト化するようにしてもよい。

【0246】これらの作成方法により作成された分野辞書201を更新する場合、PW41は、所定のタイミングでサーバにアクセスしてリストを取得し、そのリストに基づいて、分野辞書201の情報を更新する。

【0247】以上のように、分野辞書201は、様々な情報に基づいて、作成することができる。

【0248】なお、分野辞書201は、上述したように様々な方法で作成された複数の分野辞書で構成されるようにしてもよい。

【0249】その場合、文書特徴データベース作成部24が参照する分野辞書、分野辞書201に含まれる複数の分野辞書の中から、ユーザが、自分が興味あるものをダウンロードするなどして、予め選択しておくようにしてもよいし、分野辞書201が、全ての話題に含まれている全単語の内、より多くの単語を含む分野辞書を予め選択するようにしてもよいし、検索結果に基づく分野辞書を利用する場合、分野辞書201が、自分の検索語が多く登録されているものを選択するようにしてもよいし、外部のサーバにおいて、分野辞書を提供する場合、契約内容、または、利用形態等に応じて、サーバ側がユーザをグループ化し、優先するサーバ内容によって、提供する辞書を選択するようにしてもよい。

【0250】また、以上においては、分野辞書に含まれるとして記録された単語については、文書特徴データベース作成部24での重み計算時に、より重みを高く扱うように説明したが、これに限らず、例えば、文書内容処理部23が単語の重みを算出する際に、分野辞書201に含まれる単語の重みを上げるようにしてもよい。

【0251】例えば、文書内容処理部23が図3のステップS3において行う形態素解析処理において使用される形態素解析辞書の変形が可能である場合、予め、形態素解析辞書に含まれる品詞部分に新たな品詞(例えば、「一般興味品詞」等)を追加し、その品詞(一般興味品詞)を分野辞書201に含まれる単語に対応させておく。そして、文書内容処理部23が、図3のステップS4において、tf・idf法を用いて単語に重み付けを行う際に、その品詞の単語の重みを上げるようにする。このようにすることにより、文書内容処理部23が単語の重みを算出する際に、分野辞書201に含まれる単語の重

みを上げることができる。

【0252】さらに、追加された新たな品詞(一般興味品詞)が既に用意されている品詞よりも優先的に判定されるように所定の値を厳することにより、分野辞書201に含まれる単語が既に用意されている品詞である場合においても、その単語は、追加された新たな品詞(一般興味品詞)として判定されるので、その単語の重みを上げることができる。

【0253】以上のように、分野辞書201を用いることにより、ユーザが興味を示す分野の関連情報が提示されやすくなるように、データベースの単語の重みを修正することができる。

【0254】なお、以上においては、ユーザが興味を示す分野の関連情報が提示されやすくなるように、データベースの単語の重みをあげるように、分野辞書201を用いるように説明したが、これに限らず、重みを下げるために、分野辞書201を用いるようにしてもよい。

【0255】例えば、未成年のユーザで構成されるグループに対してだけ、教育上好ましくない分野の単語の重みが下がるようにするなど、特定のユーザグループに対してだけ、特定の単語の重みが下がるようにする、または、その単語の重みに「0」を乗算し、重みを「0」にするために、分野辞書201を用いるようにしてもよい。

【0256】以上においては、話題内に含まれる単語の重みを修正する場合を例にして説明したが、これに限らず、ワードプロセッサプログラム3によって作成された文書に含まれる単語にも、同様にして重みを修正することができる。その場合、話題毎ではなく、フォルダ毎あるいは1つの作成文書毎に含まれる単語の重みを修正するようにしてもよい。

【0257】以上のように、従来、ユーザのメールのやりとりにより抽出される単語に基づいて、一律にデータベースを作成してしまい、ユーザが興味を示している話題、および興味を示していない話題が同等に扱われてしまうが、図31、図34、および図36のフローチャートを用いて説明した、単語の重みの修正処理を実行することによって、ユーザがより興味を示している単語の重みを修正することができ、意義のあるデータベースを作成することができる。従って、図5の関連情報提示処理のステップS15において、ユーザの興味が高いと思われる単語を重要語として選択しやすくなることと、文書や話題の特徴ベクトルの内、ユーザの興味のある分野に関連する要素を大きくすることにより、よりユーザの興味を反映した精度の良い話題の抽出が可能となり、結果としてユーザの興味を精度良く反映した関連情報を提示できる。

【0258】また、逆に、ユーザの興味のない分野に関連する要素を小さくすることにより、ユーザにとって有用でない重要語を排除でき、無意味な特徴ベクトルの要

素に基づいた話題の抽出を抑制できるので、ユーザの所望しない関連情報の提示を抑制できる。

【0259】ところで、エージェント制御部13は、図5のステップS16の処理に対応して図7のステップS39において、文書の属性情報を選択（推薦）した理由としてデスクトップに表示させるように説明したが、その際、表示する属性情報を一部に限定するようにし、同じ属性情報より複数の表示パターンを生成することができるようにしてもよい。

【0260】図42は、同じ属性情報から複数の表示パターンを生成する例を示す図である。

【0261】例えば、図42Aに示すように、「2001年8月21日」、「Aさん」、および「パーティ」等の属性情報、並びに、「ワインとケーキ」というキーワードより選択（推薦）理由を生成してデスクトップに表示させる場合、エージェント制御部13は、図42Bに示すように、これらの全ての単語を用いて、「2001年8月21日にAさんとパーティについて話していたよ。その話題に出てきたワインとケーキに関連する情報があるよ。」のような文章の選択（推薦）理由を生成し、エージェント7の吹き出し71に表示させるようにしてもよいし、図42C乃至図42Eに示すように、上述した単語の一部を用いて、異なる文章を選択（推薦）理由として生成し、エージェント7の吹き出し73に表示させるようにしてもよい。

【0262】すなわち、エージェント制御部13は、図5のステップS16の処理に対応して図7のステップS39において、図42Cに示すように「2001年8月21日」および「Aさん」を用いて「2001年8月21日にAさんとメールで話していたよ。これに関する情報があるよ。」という文章、図42Dに示すように「Aさん」および「パーティ」を用いて「Aさんとパーティについて話していたことがあったよね。これに関する情報があるよ。」という文章、または、図42Eに示すように「Aさん」および「ワインとケーキ」を用いて「Aさんとワインとケーキについて話していたことがあったよね。関連情報があるよ。」という文章のいずれかを生成し、エージェント7の吹き出し73に表示させる。

【0263】以上のようにして、エージェント制御部13は、同じ文書を選択（推薦）理由として何度も用いた場合においても、毎回異なる文章を生成して表示させるので、ユーザの興味を低減することを抑制することができる。

【0264】なお、以上において、一度使用された組み合わせを記憶部49に記憶させておき、エージェント制御部13が選択（推薦）理由の文章生成する際にそのデータを参照し、同じ組み合わせの単語を用いないようにしてもよい。

【0265】また、推薦理由の文書生成に際して、推薦

理由を構成する単語の類義語、上位概念語、または、関連語等を、シソーラス辞書、オントロジー辞書（いずれも図示せず）等を用いて検索し、推薦理由を構成する単語として利用してもよい。例えば、図42Aの「2001年8月21日」に基づいて、「8月下旬」、「夏の終わり」等を、推薦理由を構成する単語として適用してもよいし、図42Aの「パーティ」に基づいて、「お祝い」等を、推薦理由を構成する単語として適用してもよいし、図42Aの「ワインとケーキ」に基づいて、「洋酒と洋菓子」等を、推薦理由を構成する単語として適用してもよい。

【0266】本発明のエージェントプログラム1は、上述したようにメラ2によって受信される文書やワードプロセッサプログラム3で編集される文書の他、例えば、チャット、電子ニュース、電子掲示板等の文書や音声信号をテキスト化した文書など、属性情報としてタイムスタンプが付与されている文書に対応して動作するようにさせることができる。

【0267】上述した一連の処理を実行するエージェントプログラム1は、パーソナルコンピュータに予め組み込まれるか、あるいは、記録媒体からインストールされる。

【0268】上述した一連の処理は、ハードウェアに実行させることもできるが、通常、ソフトウェアにより実行させる。一連の処理をソフトウェアにより実行させる場合には、そのソフトウェアを構成するエージェントプログラム1が、専用のハードウェアに組み込まれているコンピュータ、または、各種のプログラムをインストールすることで、各種の機能を実行することが可能、例えば汎用のパーソナルコンピュータなどに、記録媒体からインストールされる。

【0269】コンピュータにインストールされ、コンピュータによって実行可能な状態とされるプログラムを記録する記録媒体は、図2に示されるように、プログラムが記録されている磁気ディスク52（フレキシブルディスクを含む）、光ディスク53（CD-ROM（Compact Disk-Read Only Memory）、DVD（Digital Versatile Disk）を含む）、光磁気ディスク54（MD（Mini-Disk）（商標）を含む）、もしくは半導体メモリ55などよりなるパッケージメディア、または、プログラムが一時的もしくは永続的に記録されるROM42や記憶部49を構成するハードディスクなどにより構成される。記録媒体へのプログラムの記録は、必要に応じて、ルータ、モデムなどのインタフェースを介して、公衆回線網、ローカルエリアネットワーク、インターネット、デジタル衛星放送といった、有線または無線の通信媒体を利用して行われる。

【0270】なお、本明細書において、記録媒体に記録されるプログラムを記述するステップは、記載された順序に沿って時系列的に行われる処理はもちろん、必ずし

も時系列的に処理されなくとも、並列的あるいは個別に実行される処理を含むものである。

【0271】

【発明の効果】本発明の情報処理装置および方法、並びにプログラムによれば、既存の文書情報から第1の特徴語が抽出するとともに、所定の文書情報から第2の特徴語を抽出し、第1の特徴語の第1の重みおよび第2の特徴語の第2の重みを算出し、所定の条件に基づいて算出された第1の重みを修正し、第1の特徴語および第1の重みに基づいて、既存の文書情報に対応付ける関連情報を取得し、第2の特徴語に基づいて、所定の文書情報に関連する既存の文書情報を検索し、検索された既存の文書情報に対応付けられている関連情報を表示するようにしたので、ユーザの興味の度合いに応じて関連情報を提示することができる。

【図面の簡単な説明】

【図1】本発明を適用したエージェントプログラムの機能ブロックの構成例を示す図である。

【図2】エージェントプログラムをインストールして実行させるパーソナルコンピュータの構成例を示すブロック図である。

【図3】エージェントプログラムのデータベース作成処理を説明するフローチャートである。

【図4】図3のステップS5の処理を説明するための図である。

【図5】エージェントプログラムの関連情報提示処理を説明するフローチャートである。

【図6】図5のステップS15の処理を説明するための図である。

【図7】エージェントの動作等を説明するフローチャートである。

【図8】図7のステップS11の待機中の処理の詳細を説明するフローチャートである。

【図9】デスクトップ上に表示されたエージェントの表示例を示す図である。

【図10】エージェントが登場するときの表示例を示す図である。

【図11】エージェントの台詞である吹き出しの表示例を示す図である。

【図12】エージェントが待機中であるときの表示例を示す図である。

【図13】エージェントが作業中であるときの表示例を示す図である。

【図14】デスクトップ上に表示された入力ウィンドウの表示例を示す図である。

【図15】入力ウィンドウの表示例を示す図である。

【図16】デスクトップ上に表示された推薦画の表示例を示す図である。

【図17】エージェントが指示中であるときの表示例を示す図である。

【図18】デスクトップ上に表示されたスクラップ帳ウィンドウの表示例を示す図である。

【図19】エージェントが喜びの状態であるときの表示例を示す図である。

【図20】エージェントが悲しみの状態であるときの表示例を示す図である。

【図21】エージェントが横方向に移動するときの表示例を示す図である。

【図22】エージェントが縦方向に移動するときの表示例を示す図である。

【図23】エージェントが遊びの状態であるときの表示例を示す図である。

【図24】エージェントが睡眠の状態であるときの表示例を示す図である。

【図25】エージェントが立ち去るとき表示例を示す図である。

【図26】メニューボックスの表示例を示す図である。

【図27】設定画面の表示例を示す図である。

【図28】エージェントプログラムのデータベース更新処理を説明するフローチャートである。

【図29】データベースを更新させる条件を入力するユーザインタフェースの表示例を示す図である。

【図30】特徴ベクトルの例を示す図である。

【図31】第1の例の重み修正処理を説明するフローチャートである。

【図32】メールの頻度の多い順にソートされたメールの相手とその頻度を示す図である。

【図33】重みが修正された特徴ベクトルの例を示す図である。

【図34】第2の例の重み修正処理を説明するフローチャートである。

【図35】重みが修正された特徴ベクトルの例を示す図である。

【図36】第3の例の重み修正処理を説明するフローチャートである。

【図37】重みが修正された特徴ベクトルの例を示す図である。

【図38】本発明を適用したエージェントプログラムの機能ブロックの、他の構成例を示す図である。

【図39】図2の文書特徴データベース作成部による分野辞書を用いた重み修正処理について説明するフローチャートである。

【図40】図2の文書特徴データベース作成部による分野辞書を用いた重み修正処理の他の例について説明するフローチャートである。

【図41】サーバが取得した検索用キーワードに基づいて、単語をグループ化する様子を示す図である。

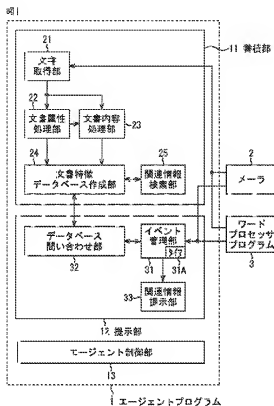
【図42】同じ属性情報から複数の表示パターンを生成する例を示す図である。

【符号の説明】

1 エージェントプログラム、 2 メーラ、 11 書籍部、 12 提示部、 13 エージェント制御部、 21 文書取得部、 22 文書属性処理部、 23 文書内容処理部、 24 文書特設データベース作成部、 25 関連情報検索部、 31 イベント管理部、 32 データベース問い合わせ部、 33 関連情報提示部、

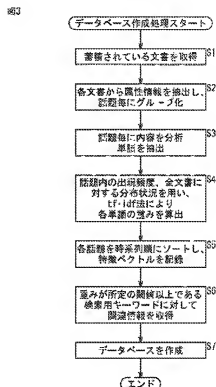
情報提示部、 52 磁気ディスク、 53 光ディスク、 54 光磁気ディスク、 55 半導体メモリ、 72 エージェント、 73 吹き出し、 74 スクラップ帳ウィンドウ、 81 入力ウィンドウ、 91 推薦皿、 201 分野辞書

【図1】



【図4】

【図3】



【図12】

図12



【図13】

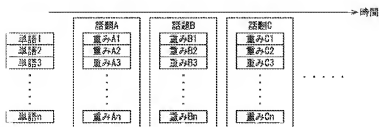
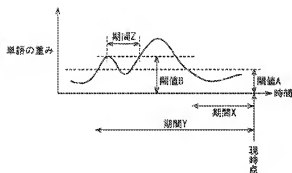


図13



【図6】



【図19】

図 19



【図24】

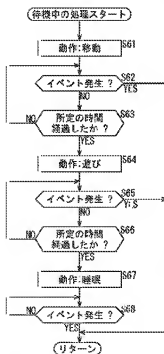
図24



【図8】

【図11】

図8



【図20】

図20



A



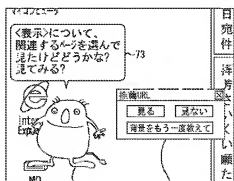
B

図11



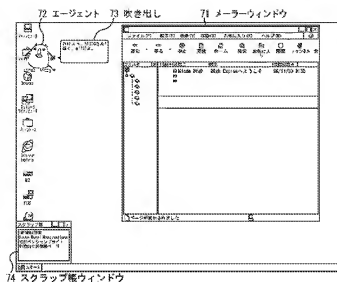
【図15】

図15



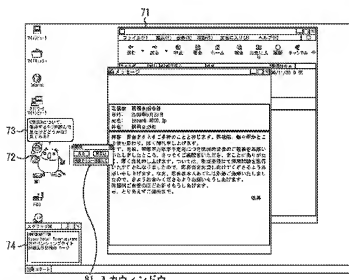
81 入カウインドウ

【図9】



74 スラップ板ウィンドウ

【図14】



81 入力ウィンドウ

【図21】

図21



A



B



C



D

【図22】

図22



A



B



C



D

E



F

E



G

【図23】

図23



A



B



A

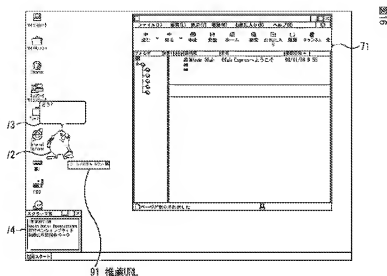
【図25】

図25

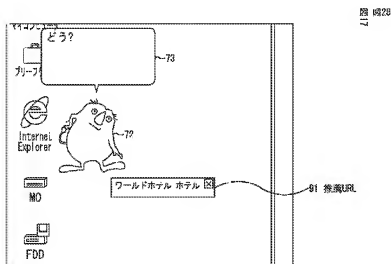


A

【図16】



【図17】

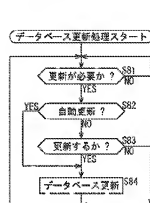


【図32】

図32

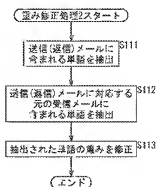
順位	強さ(回)	相手
1	35	Aさん
2	28	Cさん
3	4	Bさん
4	2	Bさん

【図28】



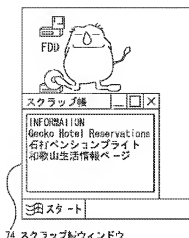
【図34】

図34



【図18】

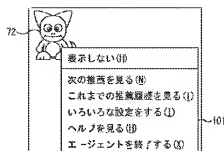
図18



74 スクラップ帳ウィンドウ

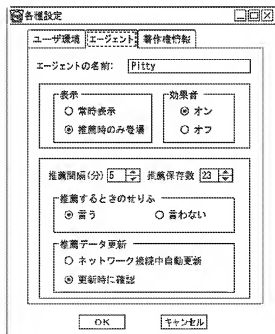
【図26】

図26



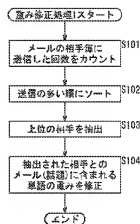
【図27】

図27



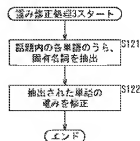
【図31】

図31



【図36】

図36



【図29】

図29

- ☐URLが ☐ 移行されたメールグループは更新必要と判断する
☐連絡URLを使い切ったメールグループは更新必要と判断する
☐最終更新から ☐ 日経過したメールグループは更新必要と判断する
☐新着メールの分析により、検索語が変更されたメールグループは更新必要と判断する

【図30】

図30

単語 →	a	b	c	d	e	f	g
話題Aの特徴ベクトル	0	0.54	0.02	0	0	0.01	0
話題Bの特徴ベクトル	0.01	0	0.41	0.33	0.16	0	0.64
話題Cの特徴ベクトル	0.51	0.35	0	0	0.48	0.37	0

【図33】

図33

単語 →	a	b	c	d	e	f	g
話題Aの特徴ベクトル	0	0.65	0.02	0	0	0.01	0
話題Bの特徴ベクトル	0.01	0	0.41	0.33	0.16	0	0.64
話題Cの特徴ベクトル	0.51	0.35	0	0	0.48	0.37	0

A

単語 →	a	b	c	d	e	f	g
話題Aの特徴ベクトル	0	0.65	0.02	0	0	0.01	0
話題Bの特徴ベクトル	0.01	0	0.40	0.33	0.16	0	0.64
話題Cの特徴ベクトル	0.51	0.42	0	0	0.48	0.44	0

B

単語 →	a	b	c	d	e	f	g
話題Aの特徴ベクトル	0	0.54	0.02	0	0	0.01	0
話題Bの特徴ベクトル	0.02	0	0.62	0.50	0.24	0	0.96
話題Cの特徴ベクトル	0.61	0.42	0	0	0.53	0.44	0

C

【図35】

図
35

単語 →	a	b	c	d	e	f	g
語類Aの特徴ベクトル	(0, 0.81, 0.03, 0, 0, 0.02, 0)						
語類Bの特徴ベクトル	(0.01, 0, 0.49, 0.33, 0.24, 0, 0.94)						
語類Cの特徴ベクトル	(0.51, 0.53, 0, 0, 0.43, 0.56, 0)						

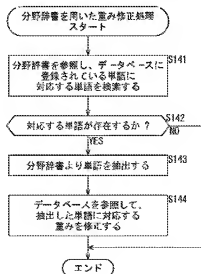
【図37】

図
37

単語 →	a	b	c	d	e	f	g
語類Aの特徴ベクトル	(0, 0.81, 0.02, 0, 0, 0.02, 0)						
語類Bの特徴ベクトル	(0.01, 0, 0.41, 0.33, 0.16, 0, 0.64)						
語類Cの特徴ベクトル	(0.51, 0.53, 0, 0, 0.43, 0.56, 0)						

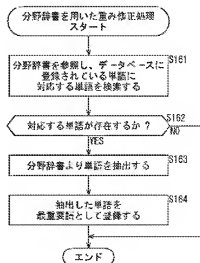
【図39】

図39

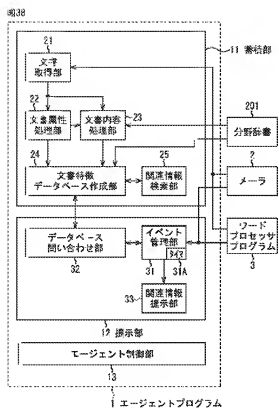


【図40】

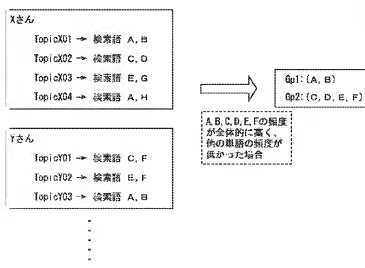
図40



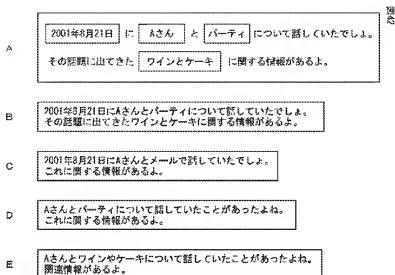
【図38】



【図41】



【図42】



フロントページの続き

Fターム(参考) 58075 ND03 NK02 NK32 NK44 NB05
PP12 PP22 PQ02 PQ32 PQ36
PQ46 PQ72 PQ74 PQ75 PR03
PR06 QM08